

The good, the bad and the ugly

Reviews of books and other creations can be highly opinionated. Editors get used to the brickbats they receive in maintaining a balance between the rights of authors and reviewers, and between fact and interpretation.

Few actresses have attracted such condemnation as did Liza Minnelli in 1972. Her portrayal of Sally Bowles in *Cabaret*, the film based on Christopher Isherwood's *Berlin Stories*, stimulated a withering response in *The New Leader*: "The film's irredeemable disaster is its Sally Bowles: changing her into an American was bad enough; into Liza Minnelli, catastrophe... Plain, ludicrously rather than pathetically plain is what Miss Minnelli is. That turnip nose overhanging a forward-gaping mouth and hastily retreating chin, that bulbous cranium with eyes as big (and as inexpressive) as saucers..."

Admit it, this makes compelling reading. Anyone who has seen the film will know what he means. But analytical and insightful it isn't. What's more, any reader would have done well to ignore the review and see the film.

Was the editor wrong to publish it? Not at all. An actress's appearance can be an essential aspect of her impact and thus merits comment that in any other circumstances would be purely bitchy. Should the editor have seen the film for him/herself before publishing? Of course not. Publication does not imply endorsement of the opinion, merely that the writer has an opinion worth hearing — if only to disagree with. In publishing reviewers' opinions, rudeness may not only be tolerated but may even be encouraged on those rare occasions when it makes its point all the more effectively. And a highly opinionated style may ensure that more readers persist to the end of the review. The trick is to ensure that there is a foundation for the review — whether it be the reviewer's reputation or analytical prose — that gives it credibility.

Like actresses, books may deserve blistering dismissal. *Nature* doesn't pursue polemic for its own sake, but has published some highly opinionated reviews in its time, and has tended to give complaining authors short shrift — they've had their say in their

books, after all. Of course, a book review must be factually accurate, give some idea of what the book is about and venture an opinion on its merits and shortcomings. But it is all the better if a reviewer takes the opportunity to expand on the broader context of the book, or provide original and complementary insights into the subject matter — even if he or she is uncomplimentary along the way.

Occasionally, an editor has to step in after publication. Correcting a factual error or publishing a reviewer's retraction is relatively easy, if painful. But sometimes the matter may be less clear cut. An example is a clarification published on page 487. On this occasion, the reviewer made a strong statement about an author's characterization of a key moment in the history of science. *Nature's* reading of the text led to a different interpretation, and we feel it worth saying so. The reviewer can explain why, based on his reading of the text and his knowledge of the history, he interpreted the text as he did, and will do so soon. Maybe others will be encouraged to comment on the history itself, as does Alvin Weinberg on page 485.

Perhaps a more efficient approach to reviews would be to encourage extreme succinctness and unanswerably pure subjectivity, blended with an element of wit. Isherwood's *Berlin Stories* again stimulated a possible model — an assessment, attributed variously to Dorothy Parker or Walter Kerr, of the play by John Van Druten based on the book, entitled *I Am a Camera*: "Me no Leica". Or even better, a review of an exhibition of nautical watercolours: "Unseeworthy".

If we were to encourage this approach, *Nature's* readers would be deprived of the objectivity and rigour they value so highly, but would have more time for their own objective analyses in the lab. Editors, for sure, would avoid much strife. Alas, such elegant brevity crosses our desks all too infrequently. ■

Fallout from EuroWars

Analogies with a current conflict seem apt.

Even the United States, with a single military command chain, inadvertently bombs Red Cross food depots. A United States of Europe for research is still off the radar screen, and collateral damage is inevitable when the interests of feuding factions prevail over coordinated strategies of mutual value. The excellent European Bioinformatics Institute came under friendly fire in 1999 when its funds dried up after member states blocked the European Commission from funding large-scale research infrastructure. Europe's commanders now risk repeating the error, hitting efforts to deepen Europe's advanced research networks (see page 475).

Support for research infrastructure has been the rallying cry of Philippe Busquin, head of research at the European Commission, with apparently whimsical support from the council of research ministers — a sort of uneasy Northern Alliance of the 15 member-state warlords of the European Union (EU) who control 95% of European research funds. Busquin wants a research policy of European unity, but warlords believe that they manage better, and are loath to yield ground to the commission, which they suspect of power-grabbing.

The third player is the European Parliament, an assembly of tribal leaders directly elected from member states, with allegiances also to pan-European political camps. A weak body, every five years it can nonetheless veto the alliance's key strategic plank — the five-year 'Framework' programmes for joint EU research. All three parties are now engaged in trench warfare over the small print of the sixth Framework, with lines and paragraphs being taken and lost in battles that are won by horse-trading and defections.

In the heat and dust, broader perspectives are sometimes lost. Busquin's request to increase funding for infrastructure to 800 million euros (US\$700 million) risks being plundered by the other parties to pay for pet topics. Warlords prefer to divide infrastructure — and its rich pickings in jobs and prestige — among themselves. The result in the past has often been ill-coordinated development, for example in neutron and light sources. Europe's fragile coalition needs to wake up to the fact that a secure mechanism for long-term planning and support of research infrastructure, at the European level, is not an item for barter, but a fundamental priority. ■





Core issues
Researchers reveal
plans to dig for
history
p476



Copy cats
Company unveils
first cloned human
embryo
p477



Check point
German terror law
puts foreign students
under the spotlight
p478



Looking up
Popular radio
telescope set for
expansion
p479

Proposed budget cuts threaten to short-circuit Grid network

Declan Butler, Paris

Computer scientists fear that plans to redraw the budget for the next five-year Framework research programme run by the European Union (EU) will cripple efforts to develop 'Grid' technology. The Grid aims to create unprecedented computing power by linking individual machines over high-speed networks using advanced software known as middleware.

Under the sixth Framework programme, which is due to run from 2002 to 2006, EU research commissioner Philippe Busquin had proposed spending 900 million euros (US\$793 million) of the programme's 17.5-billion-euro budget on infrastructure projects, including the Grid. The European Parliament has already called for changes, requesting that 350 million euros, including the money for the Grid, be transferred to the 'information society' budget, and asking for a further cut of 75 million euros to free up money for research on health, energy, transport and the environment (see *Nature* **414**, 386; 2001).

In a modified proposal, the European Commission produced a revised figure of 800 million euros for infrastructure. But Britain, Spain and Sweden are pushing for a figure of 500 million euros, with no transfer of money to other budget lines. The position of the other countries is unclear, although Italy, Austria, Finland, Portugal and Belgium seem ready to accept the 800-million figure.

Cutting the infrastructure budget to 500 million euros could place the Grid in jeopardy. Around 350 million euros of Busquin's original budget was to have been earmarked for high-speed networks and Grid projects. And whatever budget is finally agreed, 150 million euros is expected to be allocated for upgrading and running Géant — a pan-European research network that will allow Europe to keep pace with planned upgrades in the United States. That would leave projects to develop Grid software competing in the reduced pool with everything from oceanography to bioinformatics.

If so, Europe's Grid efforts would go right back to square one, says David Williams, who coordinates relations with the EU for CERN,



Data dilemma: many European facilities need Framework funds to keep abreast of high-tech advances.

the European laboratory for particle physics near Geneva: "There is a strong risk that the intricacies of the decision-making process will cause severe collateral damage."

But even if the budget is slashed, some commission officials are optimistic that funding for Grid projects may be found from other parts of the Framework programme.

CERN is relying on the EU's Grid activities to ensure that it will be able to handle the 7 petabytes (10^{15} bytes) of data that its Large Hadron Collider will generate every year when it comes online in 2006. Analysing these data will require a thousand times more computing power than CERN can deliver.

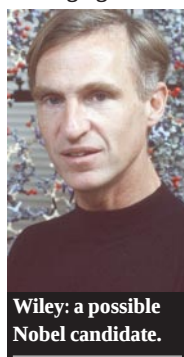
Scientists in other areas hoping for infrastructure money are similarly distressed by the signs that it is to be cut. "The impact would be disastrous," says an official at the European Bioinformatics Institute (EBI) near Cambridge. "I believe that what we do at the EBI has to be done at the European level, and if you kill that, then there just isn't any alternative funding mechanism."

A large cut to the infrastructure budget would also be a major blow to Busquin's plans to create a 'European Research Area' (see *Nature* **413**, 768–770; 2001). Support for European-level infrastructure is one of the cornerstones of the proposal. ■

Concern raised for missing biologist

Alison Abbott

Structural biologists are in shock following the disappearance of Don Wiley, one of the leading figures in the field. As *Nature* went to



Wiley: a possible Nobel candidate.

press, more than a week after Wiley's car was found abandoned near Memphis on a bridge over the Mississippi river, the FBI was still investigating.

A professor of biochemistry and biophysics at Harvard University, Wiley has been seen as a candidate for a Nobel prize. He

won a 1995 Lasker award for his resolution of the structure of the two classes of major histocompatibility proteins. These proteins bind to foreign proteins, alerting the immune system to mount a response. Wiley's work opened up new vistas in immunology.

His other key work has been on the haemagglutinin protein of the influenza virus. Wiley's analyses led to the understanding of how the virus fuses with host-cell membranes. His further work on other viruses suggests the mechanism may be general.

Wiley was last seen leaving a dinner following a meeting of the scientific advisory board of St Jude Children's Research Hospital in Memphis, where he is said to have been in good spirits. ■

Cores set to unearth hole picture of evolution

Rex Dalton, Boston

Two drilling projects to extract rock cores from formations in Africa and Australia could provide the most comprehensive analysis yet of the early evolution of life on Earth.

The two teams plan to drill nearly two kilometres into sediments that formed between 2 billion and 3 billion years ago. By analysing the resulting cores, they hope to obtain an unrivalled biogeochemical picture of a period of dramatic changes. These included a large increase in the atmospheric concentration of oxygen, which caused sulphur-metabolizing microorganisms to be superseded as the planet's dominant organisms by bacteria that respired aerobically (see *Nature* **410**, 862–864; 2001).

One team, led by Andrew Knoll, a palaeontologist at Harvard University, is expected early next year to win funding from a private foundation for its project in South Africa. The second team, which plans to drill in Western Australia, is led by geochemist Ariel Anbar of the University of Rochester in New York. It is considering applying for support from the US National Science Foundation's programme in biocomplexity, which is seeking project applications in February. There are also talks between the two teams about joining forces in the future.

"Projects like these must be done if we are



Digging deep: the projects hope for better samples than those that come from mining (right) in areas such as the Pilbara Craton in Australia (above).

to understand the 'deep time' history of this planet," says Roger Buick, an Australian palaeobiogeochemist who recently moved to the University of Washington in Seattle.

Both teams are seeking pristine rock cores — as free as possible of contaminants — from regions that offer accessible ancient rock formations. "The shortage of good core samples has become a serious drawback," explains Buick, who is a member of Anbar's



team. Previous studies have been conducted on cores taken during mining operations, which can easily become contaminated.

Knoll's team includes researchers from the Massachusetts Institute of Technology, the California Institute of Technology and Woods Hole Oceanographic Institution in Massachusetts. He declines to discuss funding details, although the initial phases are expected to cost several hundred thousand dollars. Scientists familiar with the project say the team has received strong assurances of support from the Agouron Foundation of Pasadena, California.

If this funding is secured, drilling could start before the end of next year in part of the Transvaal region of South Africa. Cores taken here could capture the geological formations of sediments ranging from tidal flats to the ocean floor laid down between 2.2 billion and 2.6 billion years ago.

Anbar's team of some 70 Earth scientists, meanwhile, is eyeing sites — such as the Warrawoona formation in Western Australia — where rocks on the surface have produced some of the earliest evidence for microbial life. Drilling here could yield cores between 2 billion and 4 billion years old. Again, Anbar says it is too early to discuss a precise price-tag for the project. Drilling could start within a year of the project winning funding.

The two projects could also provide crucial baseline data for future investigations into the biogeochemical signatures of life on other planets. "For a fraction of the cost of a space machine, we could do a nice job on samples from early Earth," says Anbar. ■

Students left cold by careers advice

Paul Smaglik, Washington

US graduate students appreciate the education that their universities provide, but feel they get too little career guidance, according to a new survey. The results, published on the Internet by the National Association of Graduate-Professional Students (NAGPS), confirm the suspicions of many university deans.

Some 32,000 graduate students responded to the survey, which was designed to measure the extent to which universities have adopted reforms proposed recently by bodies such as the National Research Council and the Association of American Universities. Such organizations argue that students should be more broadly prepared for diverse scientific careers.

The survey's data indicate that many graduate students feel that supervisors focus on training students for academic careers, sometimes to the exclusion of other options, says Kimberly Wells, past president of the NAGPS — even though academic opportunities are limited in some disciplines. Some students commented that just mentioning non-academic careers could

get them labelled "black sheep", says Wells. Others noted that they were frowned upon for considering a teaching career.

Robert Thach, dean of the graduate school of arts and sciences at Washington University in St Louis, Missouri, says he knew that his graduate school could do a better job in career guidance: "But this poll underscores my impressions dramatically." Lewis Siegel, vice-provost and dean of the graduate school at Duke University in Durham, North Carolina, agrees. "In most scientifically oriented disciplines, they've hit the nail on the head," he says.

But both Thach and Siegel criticize the survey's methodology. The total number of participating students is too low for results to be "statistically compelling", Thach says.

Most controversially, the report attempted to rank individual institutions for graduate student satisfaction — requiring only a minimum of 10 replies from a department. Deans argue that these sample sizes are much too low to give reliable results.

The NAGPS will repeat the survey next year, publishing the results in spring 2003. ■

► <http://survey.nagps.org>

First human clones get a cool response

David Adam

The announcement that researchers at a US biotech company had cloned human embryos from adult cells made headlines around the world earlier this week — and heightened the controversy surrounding ‘therapeutic’ cloning. But other researchers in the field question the claimed significance of the reported findings.

On 25 November, Advanced Cell Technology (ACT) of Worcester, Massachusetts, announced that it had created a cloned human embryo. It claimed this as an important step towards the goal of therapeutic cloning — in which cloned embryos would be used to harvest embryonic stem (ES) cells to grow replacement tissue perfectly matched to individual patients.

The ACT team fused adult cumulus cells — ovarian cells that surround eggs after ovulation — with human eggs that had been stripped of their own chromosomes. Three out of eight reconstructed eggs developed by dividing to form bundles of cells, with the best making it to six cells before stopping. Similar attempts to create embryos using skin cells met with failure.

Other groups have previously claimed to have cloned human embryos, but the ACT team is the first to publish its results (J. B. Cibelli *et al.* *J. Regen. Med.* **2**, 25–31; 2001).

In a press release, the company said that the paper provides “the first proof that reprogrammed human cells can supply

tissue”. But other experts reject this claim — and some even warn that ACT’s results indicate that human therapeutic cloning will be much harder than animal studies suggest.

Before ES cells can be isolated from human embryos, they must form a hollow, fluid-filled ball of cells called a blastocyst. The ACT clones are “nowhere near” that stage, says Alan Colman, research director of PPL Therapeutics near Edinburgh, and a member of the team that cloned Dolly the sheep.

Colman also notes that the development of ACT’s embryos compares unfavourably with experiments in animals. “Clearly, just extrapolating from the cow system into the human has not worked very well,” he says. “With cows you would expect over 30% of the reconstructed eggs to go to the blastocyst stage.”

Jose Cibelli, ACT’s vice-president of research and the paper’s lead author, agrees the work is at an early stage but argues that it is still of interest: “We understand that these are early and preliminary results, but given the importance of this emerging field of medicine we decided to publish our results now.”

While scientists debate the significance of ACT’s findings, political opponents of cloning are attempting to outlaw the research. The US Senate is considering a bill, already passed by the House of Representatives, that would ban all forms of human cloning, both reproductive and therapeutic. Observers now



Michael West, chief executive of Advanced Cell Technology, discusses the cloned embryos (inset).

expect this to be passed, and to be signed into law by President George W. Bush.

Meanwhile, emergency legislation intended to outlaw human reproductive cloning will be debated in the British parliament this week — closing a legal loophole identified by the High Court (see *Nature* **414**, 381; 2001). The court ruling effectively leaves therapeutic cloning unregulated in Britain, although the government is appealing the decision.

♦ www.liebertpub.com/ebi/default1.asp

China caught out as model shows net fall in fish

Helen Pearson

Despite a wealth of local evidence suggesting that world fish stocks are in peril, the United Nations’ Food and Agriculture Organization (FAO) has consistently reported that global catch sizes are stable or rising.

But new research suggests that the FAO statistics, which have encouraged

investment in the fishing industry, may have been distorted by exaggerated estimates of catch sizes by some countries — particularly China. The reality, say the researchers, is that fish stocks have declined alarmingly over the past decade.

The FAO classifies more than 70% of major marine fisheries as fully or over-exploited. Many populations, such as the North Atlantic cod, have already crashed. But despite these warning signs, FAO statistics for total global fish catches have increased since 1950. These seemingly anomalous figures were put down to the discovery of new stocks.

Reg Watson and Daniel Pauly of the University of British Columbia in Vancouver, Canada, have now reanalysed the FAO statistics, using information about factors such as food abundance and water depth to predict catch levels. They report the results in this issue of *Nature* (see pages 534–536). Their model mirrored the FAO figures in most regions, but China’s reported catches — which account for around 15% of the

global harvest — are twice the predicted figure. If Watson and Pauly are correct, and China has over-reported its catches, world fish stocks have actually declined by more than 10% since 1988.

“The results are stunning,” says Jane Lubchenco, a marine biologist at Oregon State University in Corvallis. “We’re on a trajectory of significant decline.” Only a drastic overhaul of fishery management can halt the global trend, she says.

Local Chinese officials, whose promotion is linked to their ability to exceed production targets in the country’s socialist economy, may be responsible for the over-reporting, believes Pauly. The central Chinese government seemed to acknowledge the problem in 1998, when it placed a cap on the figures reported to the FAO.

Dropping stocks threaten the fishing industry and world food production. Fish provides more than 15% of the world’s animal protein, and many developing countries in particular rely heavily on fish catches.



Protests as terror law targets foreigners at universities



Under a shadow: Humboldt University has been forced to supply data on foreign students to the police.

Philipp Weis and Quirin Schiermeier, Munich

Germany's student leaders and academic organizations are protesting against new anti-terrorism measures that require the country's universities to give the police information about students. They fear that the legislation will increase intolerance towards foreign students and scientists, exacerbating an already worrying problem.

The new law, introduced earlier this month, allows the police to ask universities for information on male students. The information can then be cross-checked against records from other sources, including government offices, banks, airlines, car rental companies, and the authorities that issue drivers' and pilots' licences.

Universities in the state of North Rhine-Westphalia were asked last week to supply information on 10,000 students, including their religion, place of birth and details of when they started and ended their studies. Universities in other states have also been asked to search their files.

The measures are intended to help identify 'sleepers' — terrorists who are living in Germany while awaiting an order to attack. At least three of those involved in the 11 September attacks on America had lived in Germany before entering the United States. Mohammed Atta, who piloted one of the planes that crashed into the World Trade Center, was registered as a student at the Technical University of Hamburg.

Student organizations in Berlin and Aachen have reacted by urging university heads not to comply with the law. Most universities have, however, grudgingly accepted it. The Humboldt University in Berlin is challenging the measures in court, but has already supplied the information about its students. The university intends to ask for more control over how it handles personal data, such as the right to inform

students before giving their data to the police.

Although students from European countries are also covered, most of the requests concern those from Arab nations. "It is discriminating to treat Islamic students as suspects wholesale," says Oumarou Roufaou, an informatics student from Cameroon, who represents foreign students at the Technical University of Aachen. Some 4,500 foreign science students, including many from Arab countries, are enrolled at the university.

Roufaou says that Islamic students are well integrated in campus and elsewhere, but that the legislation is making them feel intimidated. "Some have already experienced distrust or hostility, for example when looking for a flat," he says. "They are avoiding the public." Indeed, Islamic student groups declined to be interviewed for this article.

German academic organizations share the unease. "The measures are excessive and badly considered. They will hardly help track down terrorists," says Rüdiger Jütte, who is responsible for foreigners' affairs at the German Conference of University Rectors.

Manfred Osten, secretary-general of the Alexander von Humboldt Foundation, which awards grants to foreign scholars, fears the measures will increase intolerance at a time when foreign scientists in Germany are worried by increasing xenophobia and right-wing extremism, and have suffered violent attacks (see *Nature* 406, 553; 2000).

But Uwe Thomas, state secretary for education in the federal research ministry, defends the measures. "It is in the interests of all students and scientists to make sure that no terrorists are being harboured in German institutes," he says, adding that the measures are not an attempt to clamp down on scientific exchange with Islamic countries. "We are strongly interested in strengthening scientific links with the Middle East, Iran and Pakistan," he says.

Election result leaves Australian scientists fearful over funding

Peter Pockley, Sydney

Australian voters' re-election of the coalition of right-wing parties headed by Prime Minister John Howard has dashed hopes that the election might boost the country's research budgets.

Since the vote, Howard has appointed a new science minister and has shuffled certain responsibilities between departments. But it seems unlikely that he will sanction additional money for research.

Kim Beazley, the leader of the opposition Labor Party, had been campaigning on a 'Knowledge Nation' policy, promising to double spending on research (see *Nature* 414, 137; 2001). But Howard's emphasis on his leadership abilities, together with his hard-line policies on asylum-seekers, proved more attractive to voters.

During the campaign, Howard had ridiculed Beazley's policy, and commentators say Howard is unlikely to revise his opinions in favour of research.

Some Australian universities are facing severe financial difficulties, but the new minister for education and science, Brendan Nelson, who took office on 26 November, has not been given the funds to remedy the situation.

University and science leaders nonetheless welcomed the departure of Nelson's predecessor, David Kemp, who frequently clashed with the education sector.

Nelson, a former medical practitioner, honed his political negotiating skills as president of the Australian Medical Association, and has appeared to be sympathetic to the cause of research since entering parliament.

He is being assisted by a junior minister for science, Peter McGauran, who has returned to a role from which he was forced to resign four years ago over irregularities in his claims for travel costs.

Howard has also surprised many observers by moving responsibility for science and the research agencies, notably the Commonwealth Scientific and Industrial Research Organisation, from the industry portfolio to that of education.

This move has puzzled science organizations because Howard, together with the former industry minister Nick Minchin, had been pushing for research to become more commercially oriented.

Bioweapons treaty in limbo as US blocks verification plan

Geneva A conference aimed at reviewing the Biological Weapons Convention got off to an unpromising start last week, as the United States reaffirmed its opposition to attempts to equip the treaty with a verification protocol.

Negotiations to introduce a procedure for verification, which began in 1995, broke down this July in the face of US opposition (see *Nature* **412**, 365; 2001). Although a strong consensus emerged last week that more needs to be done to prevent the acquisition of biological weapons, the United States did not change its stance on verification, arguing that the surprise inspections of labs allowed under the protocol would harm US biodefence establishments and biotechnology companies.

The United States instead proposed new export controls, stronger national bioterrorism laws and a code of conduct for scientists. John Bolton, US under-secretary of state, also took the unusual step of directly accusing Iraq, North Korea, Iran, Libya and Syria of infringing the treaty. Iran, Iraq and Libya denied the accusation, and Libya added that the allegations could damage the chances of reaching a consensus at the conference.

Ombudsmen to bring order to research misconduct

Amsterdam Research ombudsmen who will investigate breaches of research ethics have been appointed by 13 Dutch universities, together with the country's Royal Academy of Sciences and the Netherlands Organization for Scientific Research, a government-funded body that distributes research funding.

The ombudsmen will tackle accusations of misdemeanours such as the manipulation of research data and plagiarism, and will report to their institution's board of governors. Although the governors will decide on the action to be taken, an appeal court, the National Body for Scientific Integrity, is also being established as part of the royal academy. All parties in a dispute will have the option of asking the court to evaluate their institution's handling of the case.

Telescope revamp gets the go-ahead

Washington Plans have been approved to expand the ability of the Very Large Array radio telescope to image distant objects. The National Science Board, the governing body of the National Science Foundation, has recommended funding of US\$58.3 million for the project over the next decade. Canada and Mexico also plan to fund the expansion.

Based in New Mexico, the instrument's 27 dish antennas make it the world's most



Farther horizons: expanding the Very Large Array will boost studies of how stars and planets formed.

powerful and widely used radio telescope. The expansion will include replacing ageing parts on the 21-year-old telescope with modern technology. A committee of the National Research Council, the operating arm of the National Academy of Sciences, last year endorsed the revamp, saying that it would be essential for future studies of star and planet formation, black holes and the origins of the modern Universe.

Patent office closes the book on microarray dispute

Munich The final stage of a long-running battle over microarray patents has been settled by the European Patent Office.

The British company Oxford Gene Technology (OGT) had been in dispute with US firm Affymetrix over the latter's GeneChip technology, used to determine which genes in a cell or tissue are switched on. OGT claimed the technology breached patents that it owned. Affymetrix and five other companies, including Abbott and Roche, had in turn challenged the validity of OGT's patent.

Affymetrix withdrew its opposition to the patent earlier this year and agreed to pay licence fees to OGT (see *Nature* **410**, 506; 2001). But opposition from the other companies continued, leading to the patent office hearing. Its ruling upholds OGT's patent but slightly limits its scope. For example, the patent applies only to microarrays made from plastic or glass. But OGT welcomed the decision, saying that it reaffirmed its position and will not in practice limit the patent.

Review finds malaise in UK computing

London Low academic salaries, outdated infrastructure and inadequate funding are eroding British research in computer science, a panel of international experts has concluded. The *International Review of UK Research in Computer Science*, sponsored by a group of governmental and computer-

science organizations, says that although the country's research is still strong, some fields are declining and others will follow.

Government funding of computer science is currently below international levels and discriminates against proposals in new areas, interdisciplinary projects and large experimental projects, the panel asserts. Two fields in particular need funds, it says — algorithmic and complexity research, which is growing in importance, and experimental computer systems research, which it says has been a strength but is now weakened.

► www.iee.org/policy/csreport/cs_report.pdf

Stamp of approval for Heisenberg's birthday

Munich The career of physicist Werner Heisenberg, who would have been 100 years old on 5 December, is being celebrated in his native Germany with the release of a commemorative stamp. This year could also see an end to the debate over the nature of a meeting between Heisenberg and Danish physicist Niels Bohr in 1941.

Heisenberg, who died in 1976, won the Nobel Prize in Physics in 1932 for his formulation of quantum mechanics. During the Second World War, he headed Germany's unsuccessful attempt to produce a nuclear bomb and it was at this time that he visited Bohr in Copenhagen.

The details of their meeting were never recorded, and questions have remained over Heisenberg's motives for visiting his former colleague. The episode forms the basis of *Copenhagen*, a play written in 1998 by Michael Frayn. The Bohr family says that by the end of the year it will release 11 documents on the meeting that Bohr either wrote or dictated to others.

► www.aip.org/history/heisenberg



The crystal ball of chaos

Is it possible to predict when nations are about to descend into internal conflict? The US Central Intelligence Agency thinks so, and has spent millions of dollars on a controversial research programme. Robert Adler reports.

The roll-call of failed states, nations in which governmental meltdowns have led to the death or displacement of millions of people, is depressingly long. From Afghanistan and Bosnia to Rwanda and Somalia, their names have become bywords for misery. And as the carnage at 'Ground Zero' in New York bears testament, the reverberations of such conflicts can spread around the world.

For decades, social and political scientists have tried to understand the factors that drive some states towards collapse, whereas others can pull back from the brink. Tools to warn of brewing instability would be extremely valuable. Yet only recently have researchers applied quantitative models to large data sets in an attempt to predict state failures. The idea has the backing of influential bodies including the US Central Intelligence Agency (CIA), but remains controversial.

Interest in applying scientific methods to the study of state collapse was spurred by Vice-President Al Gore in 1994. At his request, the CIA asked a group of academic social scientists to study the issue. This State Failure Task Force set out to identify measurable factors that could determine accurately that a nation was at high risk of failing within two years.

The researchers adopted a broad definition of state failure, including not only the worst national implosions between 1955 and 1994, but also less severe civil wars, mass killings and disruptive regime changes. That gave a workable number of cases for analysis — about three per year.

Led by Jack Goldstone of the University of



California, Davis, Daniel Esty of Yale University in Connecticut and Ted Robert Gurr of the University of Maryland in College Park, the task force built a global data set that included more than 600 variables. It amassed annual demographic, economic and environmental statistics, and developed ways to quantify concepts such as political structure and ethnic or religious divisions. The result was a database with more than two million entries.

The task force's approach has drawn fire from some academics, who characterize it as a fishing expedition that ignores the insights gained from traditional case studies and theoretical approaches. "Two million data points," quips Donald Horowitz of Duke University in Durham, North Carolina, an expert on ethnic conflict. "Obviously they had too much funding." Members of the task force respond by pointing out that decades of traditional research had yielded little more than a multitude of largely untested theories.

War and peace

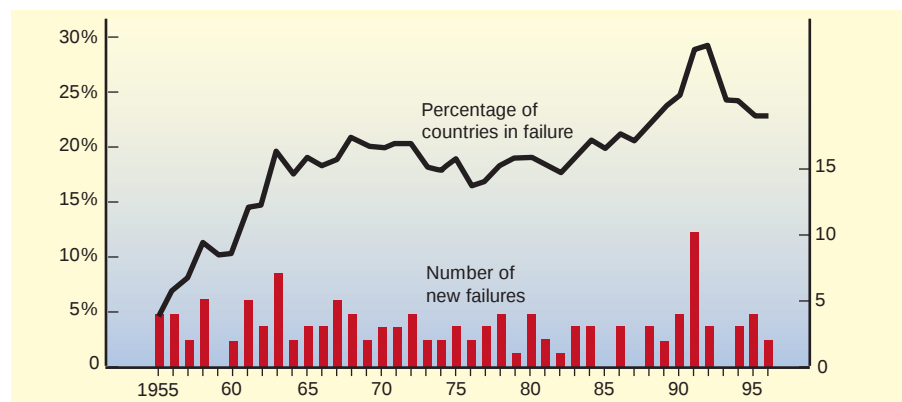
The task-force researchers matched each failure with three randomly chosen states that had avoided serious conflict, and used two methods to find patterns in the data. The first employed neural networks: computer programs that were trained over repeated cycles of analysis to look for links between the input data and the outcome of state collapse. The second method they used was multiple logistic regression, a statistical technique that considered the relationships between the different variables, determining

interactions between them and revealing which of these were most strongly associated with national failure.

These techniques threw up 13 factors for more detailed study. By 1995, the task force had developed a robust statistical model in which just three variables correctly categorized states as failed or stable two-thirds of the time, two years in advance¹. Goldstone remains surprised that a relatively simple model could even begin to address such complex events. "We had anticipated very intricate models," he says, "but we didn't need them."

The three risk factors were infant mortality, level of democracy and openness to international trade. Political theorists were not surprised by the first two. But the idea that nations imposing barriers to trade are more vulnerable to collapse is controversial. Not only have trade barriers failed to be identified as a significant factor by other experts, but the result also throws down a gauntlet to anti-globalization activists, who contend that opening a nation to the global economy exaggerates inequality and creates instability.

But one new quantitative study supports the task force's findings. Håvard Hegre, a political scientist with the World Bank in Washington DC, has tested competing theories about the factors that militate against national stability on a global data set of civil wars from 1970 to 1997 using statistical techniques to control for factors such as level of democracy, national wealth and ethnic heterogeneity². "Trade openness



Up in arms: four decades of data show a rising trend in the proportion of failed states.



Battle weary: from left, gun law in Somalia, Afghan Taliban fighters, Rwandan refugees in Tanzania.

stabilizes all kinds of regimes,” argues Hegre, “but particularly democracies.”

The State Failure Task Force continues to develop its models. For instance, it has refined its analysis of democracy to reflect ‘partial democracies’, states with a mixture of democratic and autocratic institutions — Indonesia is a prominent example. Such states are three times more likely to fail than are full democracies and pure autocracies, which were about equally stable³.

Put to the test

In its second report, published in 1998, the task force tested its refined model, based on data from 1955 to 1990, against a new data set of failed states and stable controls, with surprisingly good results³. “If we had had the model in 1991, and we predicted in real time from there to 2000, we would have had 75 to 80% accuracy,” claims Goldstone.

The task force’s third report is due to appear within the next few months. It will include new models and results concerning ethnic wars, risks from high-conflict neighbours, and state failures in predominantly Muslim countries — and so will make interesting reading for diplomats considering the prospects for post-Taliban Afghanistan and its regional neighbours. States with active Islamist groups are four times more likely to fail than are other largely Muslim countries, but openness to trade, membership of regional organizations and lower infant mortality can reduce this risk, says Goldstone.

The task force’s work is feeding into the US government’s thinking. “I’ve been

involved in probably hundreds of policy briefings across the range of government,” says Esty. “One wouldn’t want to rely on it in a rigid way, but we think it’s a very important and useful tool.”

But the task force has been taken to task by Gary King, a professor of government at Harvard University⁴. Among other flaws, King argues that the task force has overestimated the probabilities of failure for individual states by neglecting to correct for the ratio of three controls for each failure.

Some members of the task force accept the criticism, but say it makes little difference to the rankings of the states most likely to fail — which, in practical terms, is the most important result. And, despite his criticism, King lauds the task force for compiling its data set. Indeed, he has built from its work by adding three new variables — measures of the proportion of a state’s population in the military, relative population density, and legislative effectiveness — and developing a different statistical model, which he argues predicts state failures more accurately than the task force’s global model.

Other researchers are developing related models. Oxford University economist Paul Collier, currently seconded to the World Bank, studies civil wars, and has found that economic factors such as insurgents’ access to valuable natural resources, and support from abroad, are more important risk factors than the presence of an obvious grievance⁵. In unpublished work, David Laitin and James Fearon of Stanford University have found that the relative rewards to be gained by joining a rebellion are a key risk factor, more important than ethnic divisions. But Nicholas Sambanis of Yale finds that grievances can outweigh economic factors if ethnic or religious identity is threatened⁶.

A persistent criticism of all the ‘structural’ models produced by the task force and others is that they are not sufficiently responsive to fast-moving political events. “Structural analysis is not very good at telling you what will happen when,” admits Gurr.

For this reason, some researchers are trying to develop computer algorithms that can analyse news stories or other reports to warn of imminent state collapse. Pioneered by

political scientist Phil Schrodt of the University of Kansas, the algorithms analyse text to divide events into categories, essentially by asking ‘who did what to whom?’⁷.

Any approach that relies on studying reports of events, rather than primary data, risks falling foul of the ‘garbage in, garbage out’ syndrome — but the researchers involved argue that they are producing impressive results. In February this year, for instance, Craig Jenkins of Ohio State University in Columbus outlined one such model that expresses its results in terms of a state’s ‘conflict carrying capacity’ — a score from zero to 100 (ref. 8). Jenkins sees states whose scores dip below 85 for more than six months as powder kegs in which a single triggering event is likely to ignite large-scale conflict.

Early-warning system

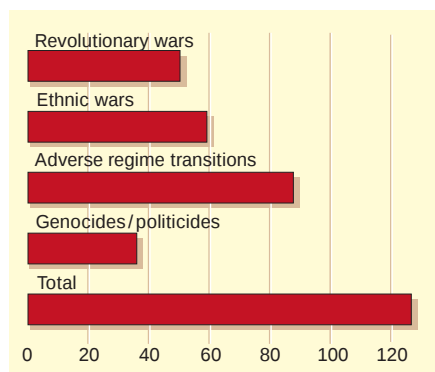
The State Failure Task Force is now experimenting with similar methods, in a project led by Barbara Harff of the US Naval Academy in Annapolis, Maryland, that aims to predict what the task force calls politicides and genocides: mass killings on political or ethnic grounds⁹. “It’s the beginning of a systematic early-warning system,” she says.

That still leaves the question of whether stable states are prepared to act on such warnings and provide assistance to prevent failing nations from sliding into chaos. Afghanistan has been a demonstrably failed state for years. But only now, following the momentous events of the past two months, is the world considering how to rebuild its shattered institutions. ■

Robert Adler is a writer in Santa Rosa, California.

1. Esty, D. C., Goldstone, J. T., Gurr, T. R., Surko, P. T. & Unger, A. N. *Working Papers: State Failure Task Force Report* (Science Applications International Corp., McLean, Virginia, 1995).
2. Hegre, H. & Gleditsch, N. P. in *Globalization and Conflict* (eds Schneider, G., Barbieri, K. & Gleditsch, N. P.) (Rowman & Littlefield, Lanham, Maryland, in the press).
3. Esty, D. C. et al. *State Failure Task Force Report: Phase II Findings* (Science Applications International Corp., McLean, Virginia, 1998).
4. King, G. & Zeng, L. *World Polit.* **53**, 623–658 (2001).
5. Collier, P. & Hoeffler, A. <http://www.worldbank.org/research/conflict/papers/greedandgrievance.htm>
6. Sambanis, N. J. *Confl. Resolut.* **45**, 259–282 (2001).
7. Schrodt, P. A. & Gerner, D. *Am. J. Polit. Sci.* **18**, 825–865 (1994).
8. Jenkins, J. C. & Bond, D. J. *Confl. Resolut.* **45**, 3–31 (2001).
9. Harff, B. & Gurr, T. R. *J. Peace Res.* **35**, 551–579 (1998).

State Failure Task Force www.bsos.umd.edu/cidcm/stfail



Power cuts: in 40 years more than 120 states collapsed, sometimes for overlapping reasons.

Goliath befriends David

Increasingly, the drugs giants are outsourcing research in drug discovery to start-up companies. Tom Clarke and Helen Pearson analyse an emerging trend, and ask what both sides expect to gain.

Eighty-five million dollars — that's how much the drugs arm of German chemicals giant Bayer agreed to pay CuraGen of New Haven, Connecticut, in January this year. In return for this upfront purchase of its stock, the biotech firm signed a deal promising to deliver around 80 obesity and diabetes drug 'targets' mined from human genome data over five years.

This move was seen by industry insiders as confirmation of a growing trend for big pharmaceutical companies to contract start-up firms to do the initial stages of drug development — the discovery of molecules that have the potential to become profitable drugs. "The size of that one deal blew everyone away," says John Hefti of Signature BioScience, a start-up in Hayward, California.

Biotech firms have always sought to sell their ideas to drugs giants. And the giants have also kept a close eye on developments in the biotech sector — on occasion buying companies that have developed a promising product portfolio. For example, in 1990 Swiss-based Roche took a controlling interest in Genentech of South San Francisco, one of the trailblazers of the biotech sector.

But over the past few years, the relationship between start-ups and larger firms has changed. Rather than cherry-picking candidate drugs developed by biotech companies, or swallowing firms that are poised to become profitable, the large firms are contracting out an increasing proportion of drug discovery — work that traditionally would have been done by chemists and pharmacologists



Big business: large firms are desperate for fresh leads to replace current blockbuster therapies.

in the firms' in-house laboratories. The companies winning these contracts include those in the biotech sector, plus others that are exploiting information technologies or novel approaches to pharmaceutical chemistry.

In part, this trend is driven by mounting economic pressure on the large companies to find replacements for existing blockbuster drugs. At the same time, drugs giants are realizing that the explosion of drug-target data from the human genome is beyond their own capabilities to pursue. But creative new ideas for exploiting such data are feeding numerous start-up companies — and the big firms want to buy into a piece of this action.

Discovery channels

Some within the industry believe that a fundamental shift is under way. Ultimately, they predict that the discovery of new molecular leads may become the preserve of smaller companies, leaving the giants to develop candidate molecules into practical drugs, and putting these through clinical trials. "It's certainly a plausible scenario," says David Floyd, vice-president of discovery chemistry for the US company Bristol-Myers Squibb.

Others argue that the current vogue for outsourcing drug discovery is a sensible



strategy while uncertainty reigns over which technologies will be needed to reap a windfall from genomics. Once clear technological winners emerge, these experts contend, the big firms will take the outsourced work back into their own labs.

In any case, argues Ben Shapiro, executive vice-president for worldwide licensing and external research with the US giant Merck, the big companies will need in-house expertise in drug discovery to inform decisions on which start-ups to link with. "The better the internal scientists, the better our decisions about external relationships," he says.

What is clear is that the giants' existing sources of income are drying up. Many of the highest-earning drugs will soon be coming off patent. And the drugs 'pipeline' of compounds awaiting development and testing is running dry. To maintain profits, the big firms need novel drug leads, fast. Start-ups are now getting bigger and better deals with the giants mainly because the latter are "desperate for the drugs", says Joseph Dougherty, a biotech analyst at Lehman Brothers in New York.

So what sorts of technologies are featuring in the current wave of contracts? Rapid and sophisticated versions of structure-based drug design, in which molecules are



Turning genomics into drugs: researchers at De Novo (above) and Ray Stevens of Syrrx (right).

synthesized to fit known sites on proteins involved in disease, are proving popular. Astex Technology, for instance, is a company in Cambridge, UK, that uses X-ray crystallography to determine the structures of drug targets. It then selects fragments of potential drug compounds that look as if they should bind to the active site under investigation. Astex's scientists let these molecules bind to the target site and they then determine the structure of the resulting complex.

This allows Astex to look for weak interactions between the potential drug molecules and their targets that might have been missed using traditional bioassays. Molecules that show promise can then be tweaked to enhance binding. "We're doing intelligent chemistry," says Harren Jhoti, one of the company's founders. The two-year-old firm already has a deal with the Anglo-Swedish firm AstraZeneca to provide crystal structures of cytochrome P450s — enzymes that metabolize drugs — complexed to AstraZeneca's compounds.

Working out which of the estimated 10^{60} potential drug candidates might bind to a protein's active site is a formidable task. And several start-ups are now offering bespoke software to help design molecules with just the right structure. "A computer can generate a far larger number of solutions than a chemist," says David Bailey, chief executive of De Novo Pharmaceuticals in Cambridge, UK.

De Novo, a 1999 spin-off from the University of Cambridge's pharmacology department, feeds known structures of protein binding sites into software that constructs virtual molecules piece-by-piece to fit. To avoid making the mistakes of early structure-based drug designers — whose tailored molecules often proved difficult to produce — each time a fragment is

added, the structure is checked for ease of synthesis and similarities to molecules that are known to be toxic. Last year, the company struck a deal — the details of which have not been disclosed — to provide drug leads to the Franco-German company Aventis.

Binding contract

But Raymond Salemm, chief scientific officer of 3-Dimensional Pharmaceuticals (3DP) in Yardley, Philadelphia, argues that approaches that rely on establishing the structures of potential drug targets are unnecessarily cumbersome. Instead, 3DP is probing changes in the thermodynamics of the bonds between a library of 200,000 drug-like molecules and putative drug targets, before details about the structure of the binding site are known. Essentially, 3DP looks for changes in the temperature at which the protein targets 'denature', or break apart, which increases when molecules are bound tightly to them.

When they spot a promising drug candidate, 3DP's chemists produce a range of related molecules and screen them through iterative rounds of molecular fine-tuning. "If you have a target, we can develop a high-affinity compound," claims Salemm. The company is already doing just that for Bristol-Myers Squibb. In return for producing molecules that bind with high affinity to at least 30 targets, 3DP is getting an upfront fee of \$19 million, plus research funding of \$14.4 million over the first three years of the collaboration, and a share in subsequent royalties.

Other companies are developing novel high-throughput screens in which thousands of compounds can be tested for biological activity. Signature BioScience, for instance, throws compounds onto a range of cell lines and, by firing radio- and microwaves at them, monitors any physiological effects, such as a change in shape, ion flow or protein movement. "It's quick and dirty," admits Hefti, but by rapidly weeding out compounds that generate unwanted effects the company hopes to home in on potential drugs, again without knowing the target beforehand.

In most of the deals currently being



struck, the start-ups get a fee up front and the rest on delivery. The drugs giants "don't pay for much unless we're successful", says Hefti. This allows the start-ups to recoup the investment made in them by venture capitalists. And, in the most part, the start-ups are happy to let the larger firms take over the later stages of drug development: they accept that they lack the financial resources, or familiarity with the regulatory procedures, needed to proceed through preclinical drug development and clinical trials.

Dreams of grandeur

But some of the smaller companies have ambitions to join the ranks of the giants. Vertex Pharmaceuticals of Cambridge, Massachusetts, is one such firm. In May last year, it struck a deal with Swiss-based Novartis promising to deliver eight compounds targeted against protein kinases, a family of enzymes implicated in a wide range of diseases including cancer and cardiovascular disease. Vertex will be responsible for drug development and the first phase of clinical trials — in which drugs are tested on small numbers of patients to check for side effects — before handing over to Novartis. If Vertex meets all of the milestones agreed under the deal, it could earn \$800 million.

How the relationship between start-ups and the big firms will develop in the long term is unclear. But given the diversity of drug discovery technologies that are currently emerging, the trend for outsourcing seems set to continue for some time yet.

What the industry is waiting for is a proven technological platform for translating genomics data into candidate drugs on a factory scale. "Some day someone's going to become the Henry Ford of drug discovery," predicts Ray Stevens of the Scripps Research Institute in La Jolla, California, and co-founder of Syrrx, a company that is working on structure-based drug design. Stevens, for one, believes this breakthrough will come from a fleet-footed start-up, rather than from a large pharmaceutical firm. ■

Tom Clarke and Helen Pearson are both members of *Nature's* science writing team.



On target: Merck uses in-house expertise to help decisions on which projects to outsource.

An eye on the future

The next president of Germany's Max Planck Society is putting aside a glittering research career in developmental biology to wrestle with politics, ethics and budgets, says Alison Abbott.

Scientifically, Peter Gruss seems to have it all. He is the most highly cited developmental biologist in Germany, and just two years ago he launched a groundbreaking project to open up the black box of brain development.

So when he was approached by the committee looking for a successor to Max Planck Society (MPS) president Hubert Markl, it must have been tempting to politely decline. The political climate for research in Germany is stormy — the issue of human embryonic stem (ES) cells is the focus of widespread public hostility — and the current economic downturn means that budgetary expansion is out of the question.

Gruss admits he was torn: "I was surprised when I was approached, and it took me at least a week to consider the position, since I am at such a critical point in my research." But accept he did, and last week his nomination to take over from Markl as head of Germany's premier research organization was confirmed by the MPS's governing senate.

Gruss, who runs a 30-strong lab at the Max Planck Institute for Biophysical Chemistry in Göttingen, is best known for his work on the Pax genes in mice. His group, for example, identified the central role of a gene called *Pax6* in the development of the eyes and forebrain. Two years ago, having just turned 50, he launched a major project to systematically determine the genetic basis of the development of the cerebral cortex, about which little is known. Nearly half of the researchers in his lab are now busy developing the wide range of functional genomic techniques that will be run in parallel to crack the problem.

He is certainly not the archetypal MPS president. Markl had already served as head of the DFG, Germany's main agency for research grants, before taking over as the society's president in 1996, and was well known as a political animal. Gruss, by contrast, is the scientist's scientist — informal, approachable



Different worlds: Peter Gruss will be leaving the safe haven of his laboratory for the politics of the Max Planck Society headquarters in Munich (right).

and happiest shooting the breeze with his fellow researchers.

So does Gruss have the political skills to succeed in one of the most important and high-profile jobs in German science? One positive sign is his engagement in recent months in Germany's contentious debate over human ES-cell research. Gruss has repeatedly spoken out on the importance of this work, without which it may not be possible to develop therapies based on adult stem cells — touted by opponents of ES-cell research as a viable alternative.

This message still needs reinforcing, says Gruss. "Many people believe that scientists will demand a continuous supply of surplus embryos, whereas what we really want are a few good cell lines to help us learn how they differentiate and therefore how to reverse differentiation — that is, how to reprogramme adult stem cells."

Gruss admits to a personal interest in the debate. Two years ago he founded the company DeveloGen to exploit his group's discovery that one of the mouse Pax genes causes adult pancreatic stem cells to differentiate into insulin-producing cells. He is excited about the therapeutic potential in diabetes, and is concerned that restrictions on ES-cell research will hold back the entire field. "I am defending my research area," he says.

But he accepts that it may not be easy to be so outspoken as MPS president. "I will have to consider the views of other members of the MPS when I am speaking on behalf of the

whole society."

Gruss will take over the reins from Markl in June, and acknowledges that he has a tough act to follow. Markl has made significant headway in tackling long-standing problems such as a lack of flexibility in adopting new areas of research and an absence of women in top positions. He also handled the unpleasant business of closing institutes in western Germany so that new institutes in the east could flourish.

Gruss plans to continue Markl's initiatives to make the MPS more dynamic. He will also face the challenge of replacing nearly a third of the MPS's 260 research directors during his six-year tenure, as the generation of baby-boomers comes to retirement.

What Gruss would really love to achieve is a major increase in funding for the MPS. "We need to make it clear to the public that basic research establishes the basis for the future of society, and we don't just do it for fun," he says. "We need to be internationally competitive. We must think in terms of budget increases much higher than in the past."

Then, checking an enthusiasm that might be interpreted as naivety, he adds: "But I am very pragmatic and fully aware that it has to be seen in the context of general economic growth — one cannot demand the impossible." It seems that Gruss is already adapting to the political realities of his new role. ■

Alison Abbott is *Nature's* senior European correspondent.



ALISON ABBOTT

Basic research establishes the basis for the future. We don't just do it for fun.

Japan's funding cuts hit the future of science

Raising the cost of postgraduate education is likely to exclude many promising students.

Sir — Since Prime Minister Junichiro Koizumi initiated structural reforms, the research system in Japan has prepared for extensive and rapid changes (see *Nature* **412**, 364; 2001). In August, Koizumi's cabinet proposed a reduction in the national scholarship fund managed by the Japan Scholarship Foundation (JSF). The main components of the plan, to be implemented by 2005, are a reduction in interest-free loans, transfer of low-interest (3% per year) loans to the private sector, and abolition of repayment exemptions for scientists in the public research sector. Last month, abolition of JSF itself was added to this plan. In the budget plan for fiscal year 2002, there is a proposal to cut the number of interest-free loans and to reduce the total JSF budget by 9.9%.

In 2000, one-third of the graduate student population (75,290) received JSF scholarships and 66% of scholarship students received interest-free loans. The new scheme will hamper the ability of students to cover living expenses because most scholarship students will have to repay loans with interest, which could reach more than ¥6 million (US\$50,000) by the time they leave graduate school.

Like the merger of Japan's three main space programmes and research institutions (see *Nature* **412**, 843; 2001), the new proposal has shocked researchers and students.

The national scholarship system in Japan is poor compared with those in the United States and the United Kingdom. For example, there are few private scholarships, and the average JSF scholarship (about ¥1.2 million or US\$10,000 a year) does not cover tuition and living expenses. A graduate science student on a stipend could expect roughly US\$15,000–20,000 per year plus payment of tuition fees in the United States or £10,000 (US\$15,000) a year in the United Kingdom — though of course there is competition for these. But in Japan, most students work part time and/or depend on their parents' financial support to supplement meagre government support.

Even so, JSF scholarships are the major source of funding for graduate students. One institute that aids promising young researchers, the Japan Society for the Promotion of Science, is increasing its numbers of PhD fellowships. But currently, only 5% of PhD students are recipients. In June, the government announced a plan to improve the stipend system, winning approval from graduate students. Not only has this plan not yet

taken shape, however, but a reduction is already being implemented.

A national scholarship scheme that burdens graduate students will impede the progress of research in Japan. I foresee a time when only the rich can attend graduate school and many promising students will be excluded. Not only will students lack opportunity, but our country will be unable to maintain the quality of its research. If Japan wants to maintain and

even advance its status in science and technology, it should expand rather than reduce financial aid for graduate study. The government should remember that although national investment in education and research shows no short-term results, it is a wise long-term investment.

Eisuke Enoki

School of Medicine, Faculty of Medicine, Kobe University, Kusunokicho 7-5-1, Chuo-ku, Kobe, Hyogo 650-0017, Japan

Could sale of fossils be the key to ending theft?

Sir — Although the federal legislation being prepared to safeguard the fossil record may attain some of the wished-for goals described in your News story "US lays out bare bones of fossil protection package" (see *Nature* **413**, 555; 2001), experience and theory suggest that fossil theft from federal lands will continue.

Assuming there is strict enforcement, criminalizing the unauthorized removal of fossils and increasing penalties may reduce pillaging in the short run, but over the long term the trade in looted archaeological and other cultural property has increased since efforts started being made to control it.

Black markets for palaeontological specimens continue to flourish in China (see *Nature* **406**, 930–932; 2000) despite central government efforts to crack down on fossil smuggling. Colin Renfrew, professor of archaeology at the University of Cambridge, has concluded that 30 years after the adoption of the 1970 UNESCO convention to halt the illicit trade in cultural property, the destruction of archaeological sites through looting has increased rather than diminished (see *Trade in Illicit Antiquities*, eds N. Brodie *et al.*, xi–xii; McDonald Institute for Archaeological Research, Cambridge, 2001).

Price theory suggests that a divergence between a market price and an officially imposed price (because the legal sale of fossil specimens on federal lands is prohibited, the price is close to zero) fosters a black market and creates incentives for theft and looting, resulting in a permanent loss to our knowledge of the fossil record. In other sectors of society, creating a market and a price for a resource that had previously been cost-free, such as clean air, has resulted in conservation of the good, in this instance reflected in a

sharp fall in air pollution from coal-fired power plants in the American Midwest.

Applying the profit mechanism and market incentives to the management of palaeontological resources could similarly result in the improved conservation and preservation of the fossil record.

Edward Krowitz

98 Riverside Drive, Apt. 7E, New York, New York 10024, USA

Messenger RNA: origins of a discovery

Sir — In his review of James Watson's *Genes, Girls and Gamow* (see *Nature* **413**, 775–776; clarification *Nature* **414**, 487; 2001), Horace Judson attributes the discovery of messenger RNA to François Jacob, Sydney Brenner and Matthew Meselson.

In fact, Jacob, Brenner and Francis Crick, at an informal meeting on Good Friday 1960, suddenly 'discovered' the unique RNA found first in 1956 by Elliot Volkin and Lazarus Astrachan. Good accounts of this event can be found in *The Statue Within* by Jacob and *What Mad Pursuit* by Crick.

In several publications in 1958, Volkin and Astrachan thoroughly described the unusual properties of this RNA, which they termed DNA-like RNA. These were precisely the properties that Jacob and Jacques Monod sought to assign to the unstable intermediate (which they called X), necessary for the synthesis of galactosidase.

Out of that Good Friday discussion on the lactose operon came the realization that Volkin and Astrachan's DNA-like RNA was indeed the genetic messenger, hence messenger RNA (mRNA).

Alvin M. Weinberg

Oak Ridge Associated Universities, PO Box 117, Oak Ridge, Tennessee 37831-0117, USA

Global energy prospects

Choices, challenges and uncertainties for the not-so-distant future.

KARIM MOHSEN/GETTY IMAGES

Hubbert's Peak: The Impending World Oil Shortage

by Kenneth S. Deffeyes

Princeton University Press: 2001. 285 pp.
\$24.95, £17.95

Megawatts and Megatons: A Turning Point in the Nuclear Age?

by Richard L. Garwin & Georges Charpak
Knopf: 2001. 384 pp. \$30, £20.77

Tomorrow's Energy: Hydrogen, Fuel Cells and the Prospects for a Cleaner Planet

by Peter Hoffmann

MIT Press: 2001. 320 pp. \$32.95, £22.50

Stuart Young

We have heard the words 'energy crisis' repeatedly over the past 30 years, and, looking at the Middle East now, we can expect to hear them again. Yet the crisis, if it comes at all, will not arise from the drying-up of one production source or another, but will be caused by us, the energy consumers, if we do not make timely and efficient use of the alternatives to conventional sources such as oil. These three books carry this message about the (not so distant) future of global energy, although they do so from different viewpoints, dealing with oil, nuclear energy and hydrogen, respectively.

Hubbert's Peak contains some very practical information about the origin of oil, exploration for it and its production. On these matters, the geologist author Kenneth Deffeyes has written a most readable handbook which is well illustrated and has copious notes. But his book is more than that. We are introduced to the Hubbert of the title, a geologist who, in 1956, despite the scorn of others in the industry, predicted that US oil production would reach its peak, and then decline, soon after 1970. It did. Deffeyes explains how he has adapted Hubbert's statistical approach, with its elements of fact and hunch, to predict the turning-point in world oil production. Not, that is, the point when total oil reserves will be exhausted, but merely the first sign of a permanent downward trend. And, he expects, the beginning of consternation.

When will that be? About 2005 — possibly before — but not by any stretch of the figures as late as the end of this decade. So, if he is right we have, at most, two or three years in which to prepare for yet another price shock, and to accelerate our move away from oil as fuel. The strength of the book lies in its solid background and well-explained basis for that single prediction. Deffeyes has a light-hearted way of dealing with weighty matters, and his narrative is enlivened, although sometimes dislocated, by personal anecdotes. He completes his survey by mentioning some of



An end to oil refineries? Calculations suggest a permanent downward trend in oil reserves by 2005.

the other energy sources that the world will be driven to use, but does so with less vigour than he applies to those matters directly concerning oil. The book will best be used as a practical and colourful adjunct to technical and economic studies of oil production.

Replacement of oil as a fuel for electricity generation will not be regretted. Crude oil is too valuable a commodity to be burnt on so large a scale, and by doing so we also add significantly to the global greenhouse effect. A return to greater reliance on coal seems inevitable, but only in the short term, for it faces even sterner criticism as a generator of greenhouse gases than does oil.

Nuclear energy almost completely avoids this main environmental danger of fossil fuels, but carries its own burden of risks and liabilities. Foremost is the perceived link between nuclear power generation and nuclear weapons. For many years, defenders of nuclear power have told us that this link is imagined or not relevant in practice, yet the widely held fear of nuclear power, extended to radiation hazards generally, has not gone away. We now have a book, *Megawatts and Megatons*, in which all aspects of the interaction between these two applications of nuclear energy are acknowledged and frankly examined, and the awesome global implications laid bare. The joint authors, Richard



Black alternative: if oil supplies dry up, there will be a short-term return to greater reliance on coal.

Garwin and Georges Charpak, are distinguished physicists who are highly qualified for this task. *Megawatts and Megatons* is largely based on their French-language work *Feux Follets et Champignons Nucléaires* (Odile Jacob, 1997) but includes some current allusions and data.

Garwin and Charpak cover a broad and dauntingly complex field, but the sequence of topics is logical and the text is carefully signposted throughout. In this jungle you

clarification *Nature's* review by Horace Freeland Judson of James D. Watson's book *Genes, Girls and Gamow* (*Nature* 413, 775–776; 2001) contains the statement, "Watson arrogates chief credit for messenger RNA to work he did at Harvard...". The reviewer stands by his comments. The reviewer has a right to his interpretation of Watson's account of his contribution to the discovery of mRNA, but *Nature* does not agree with, and dissociates itself from, the reviewer's statement. We regret the distress that this has caused. Editor, *Nature*.

HUTTON-DEUTSCH COLLECTION/CORBIS

know just where you are. The prose is faultless and a pleasure to read, even though the book is long and the amount of detail embodied in the text is vast. An admirable guiding principle can be discerned throughout — that the concerned reader should not be denied any detail, however small, if it might possibly affect their judgement on the contentious issues before them. This inevitably leads to some difficult passages, with which the reader with no knowledge of basic nuclear physics may struggle. Diehard opponents of nuclear power will no doubt say that, in proposing a safe path to its future use, Garwin and Charpak are partisan. No careful reader of this book could support that claim. The compilation is a full and fair contribution to understanding, and should be studied by everyone concerned with the problems of nuclear power.

Peter Hoffmann writes about the future for hydrogen and fuel cells in *Tomorrow's Energy*. As editor and publisher of *The Hydrogen & Fuel Cell Letter*, he has information on relevant development projects worldwide at his fingertips. In the eye of an enthusiast, hydrogen is the ideal fuel, for whether it is burnt in an engine for propulsion, or used in a fuel cell to generate electricity, the only emission is water. Yet the wider view reveals uncertainties, among them the question of effective generation of the gas. Hydrogen should perhaps be seen as a medium for transferring and storing energy, rather than a primary source.

Hoffmann's book is rich in references to small-scale developments, but poor in data presentation. In more than 250 pages of solid text there is not a single diagram, table or graph. Yet there are two dozen photographs, mainly of experimental vehicles, which add nothing to the technical understanding of the ventures they advertise. Science-based readers will feel deprived. ■

Stuart Young is at Box 4, Noordhoek 7985, South Africa.

Somewhere over the rainbow

The Rainbow Bridge: Rainbows in Art, Myth, and Science

by Raymond L. Lee Jr & Alistair B. Fraser
 Pennsylvania State University Press: 2001.
 408 pp. \$65

Philip Ball

Faced with John Keats complaining about all charms flying at the touch of cold philosophy, we might be inclined to respond, "Oh, not that again!" But this book by meteorologists Raymond Lee and Alistair Fraser shows just how crude, prosaic and clumsy the art/science debate is apt to become, because the book is so much the opposite. Stunningly well informed about the art, science, philosophy and history of all eras since the Periclean Golden Age, unerringly elegant, flatteringly intelligent and beautifully illustrated, it is a masterful piece of accessible scholarship.

The authors have, of course, the perfect subject for bringing together not only art and science but myth, nature and anthropology. And Lee and Fraser refuse to peddle the simplistic device that celebrates Newton's matter-of-fact 'unweaving' of the rainbow and then shows how artists persisted in getting the rainbow wrong. Rather, we see how both art and science are represented by a multitude of voices, making their interplay more rich and complex than is commonly portrayed. I would not have expected the poet Wordsworth to spring to Newton's defence, for example, but here he is: "The beauty in form of a plant or an animal is not made less but more apparent as a whole by more accurate insight into its constituent properties and powers." The physicist Richard Feynman was unable to put it better over a century later.

And it is not hard to sympathize with the

American painter Frederic Edwin Church in 1883: "I wish science would take a holiday for ten years so I could catch up." The book is a joy because of such things, whether you have ever marvelled at a rainbow or not.

But who has not? The wonderful insight that emerges from the book is that, although we all imagine we know just what a rainbow looks like, painters of realistic landscapes reveal an astonishing variety of ways in which the rainbow is perceived. Rubens' version in *Rainbow Landscape* (c.1636) is way off beam, the painter making the classic mistake of representing it as a solid object that can be seen obliquely. The rainbow always faces the viewer in the plane, and moves when we move.

John Constable studied atmospheric phenomena in pedantic detail, yet his rainbow arching over *Salisbury Cathedral from the Meadows* (c.1831) is impossible, because the sunbeams show that the Sun is too high in the sky for the rainbow to be visible at all. In his defence, Constable was not averse to artistic licence, and probably considered it more important here to place the symbol of optimism over the storm-threatened church.

Impossible sunbeams also undermine Eric Sloane's rainbow in *Earth Flight Environment* (1976), for the beams should always be radii to the arch of the rainbow. We needn't get too indignant about Caspar David Friedrich's bizarre achromatic eyebrow rainbow from c.1810, apparently gracing a night-time sky, for naturalism was probably never of much significance for this supremely Romantic painter, and a devotee of Goethe was unlikely to honour Newton's spectrum.

Lee and Fraser miss no opportunity to explore the rainbow's many subtleties, giving us plenty of colour theory, wave optics and cloud microphysics. They also present a thoughtful survey of the rainbow as an advertising icon. By placing a lottery's 'pot of gold' at the end of a rainbow, the advertisers are inadvertently reminding us just how unattainable it is. Of the rainbows of popular culture, the one most sadly omitted here is the magical arch that symbolically conjures a Technicolor Oz from a monochrome Kansas.

Rainbows are genuine miracles because they reveal, for a fleeting moment and in a structure that seems a mile high, one of nature's best-kept secrets — what light contains, the origin of colour. And the rainbow is truly a bridge, not just between art and science but between myth and reality, heaven and earth. Classical commentators such as Cicero were torn between explaining the rainbow as a natural phenomenon and celebrating it as an emblem of the gods. The Bifrost bridge spanned Midgard and Asgard in Norse mythology, and its shattering was a suitably dreadful image to herald the Twilight of the Gods. Not all cultures revered the rainbow: some considered it an evil omen, and that is surely what it looks like in Dürer's *Melencolia I* (1514), framing a fateful comet. To point at



Rainbow tunnel: a colourful welcome to a tunnel in Marin County, California.

the rainbow was to break an awful taboo in cultures from Mexico to Hungary.

This is a substantial book in all senses: at 400 or so large-format pages on dense paper, it is not something to tuck in your briefcase. Nor does its attention to detail make for light reading. But those things do nothing to alter the fact that we need more books like this, unafraid to assume a degree of commitment and cultural understanding in the reader without ever losing clarity and accessibility. It will bring some colour into your life. ■

Philip Ball is a consultant editor for Nature. His latest book, Bright Earth: The Invention of Colour, is published by Penguin.

A severed thread

**Medicine and the German Jews:
A History**

by John M. Efron

Yale University Press: 2001. 343 pp. £35, £27.50

John Galloway

John Efron tells us that in today's Germany, out of 200,000 doctors only about 300 are Jews, the same number as when the Second World War began in 1939. Yet in 1933 there had been 5,500. Although it could be difficult to be neutral about such facts, Efron states them objectively, and says why this situation occurred and what it meant — and means. We learn hard lessons about the historical (and the present-day) practice of medicine in Germany — as elsewhere.

The destruction of the Jewish doctors was both an integral part of, and a sub-plot in, the Nazi regime's policy of exterminating entire national and international cultures. In Britain and elsewhere, we tend to think of the Holocaust as part of the Nazi war effort during the Second World War. Efron shows that, on the contrary, the policy of extermination was independent of the war, except that conquest brought more people within its reach. Indeed, it can be regarded as the final act of religious and secular conflict dating back to the Middle Ages.

Efron describes a wonderful ambiguity in the attitudes of the religious and secular authorities to Jewish doctors from the thirteenth to the fifteenth centuries. Royalty, the aristocracy and the clergy used the services of Jewish doctors widely and publicly. Yet at the same time, both church and state forbade Christians to be treated by Jews. That they repeated the ban at intervals suggests that the population ignored it. Indeed, in the thirteenth century, although Jews accounted for no more than 1% of the population of Europe, in many areas half the doctors were Jewish, clear witness to their popularity.

Jewish doctors were highly regarded, but



Lily pads of Wisconsin

A Sand County Almanac by Aldo Leopold, one of America's pioneering conservationists, was first published in 1949. It has now been republished (Oxford University Press, \$35) with photographs

by Michael Sewell. The book celebrates the changing seasons on Leopold's Wisconsin farm, and is said to be one of the most influential works on humans and the environment.

for different reasons at different times. Unlike the Greeks, there does not seem to have been an identifiable Jewish medical tradition, but their ability to translate medical books gave them access not only to the Greek tradition but also to Arabic and Indian knowledge. The very fact that they were Jewish also seems to have implied a familiarity with ancient knowledge and practices not available to others. This assumption cut both ways — it could be trotted out as secret and sinister when the need arose, and there was a tradition of accusing Jewish doctors of malfeasance when it seemed convenient. In 1348, foreshadowing the Holocaust, all Jewish doctors in Germany were burned alive, having been found guilty of 'causing' the Black Death.

The book also traces a complementary theme of Jews not as doctors, but as patients. Gradually, the idea grew that the Jews were not as healthy as the rest of us, despite evidence to the contrary. Some early modern Jewish doctors made a special study of Jewish diseases, which may have heightened the sense of a 'racial' difference. The Nazis used this 'genetic science' to justify their policy of racial purification to attain their goal of strengthening the nation.

Jews were early enthusiasts of a scientific approach to medicine — apparently not something that endeared them to their non-Jewish colleagues. In the nineteenth century they were accused of a mechanistic and reductionist approach to the human body and its ills, rather than the more holistic approach generally favoured by doctors. In time, this led to the rise of medical specialisms, with the specialists being predominantly Jewish. This had an unfortunate

consequence. In the First World War, when most of the German medical workforce was drafted, Jewish specialist doctors were not seen to be useful to the war effort — too effete and esoteric to treat the victims of high explosives and machine-guns. They became tarred with the brush that blamed Germany's defeat on the Jews.

In the nineteenth century there had been another issue — too many doctors. The medical profession expected status and a good income. In the period from 1889 to 1898, when the German population grew by about 11%, the number of doctors rose by nearly 60%. Jewish doctors and dentists in German and Austrian cities were in the majority, and increased competition for patients resulted in a loss of income and then status. The medical profession looked for a remedy and turned to, and then on, the Jews, particularly immigrants from Eastern Europe whom they claimed were unfit to be doctors. Some 'assimilated' Jews (including Sigmund Freud) supported this claim at first. Anti-Semitism exploded at every level in the medical schools.

When the Nazis came to power the different historical strands finally came together in a lethal intertwining. German non-Jewish doctors saw the idea of 'national hygiene' — the purging of the national 'body' of the genetic traits that weakened it — as an opportunity for advancement and power. They readily made the first fatal step of rallying to the siren call to leadership. How easily leaders lose their humanity. They slipped from a concern for people's health to the abstract and dangerous idea of racial health — and its lack. Like many others in German public life,

doctors accepted the opportunities for advancement that the removal of their Jewish colleagues afforded. Pledging the medical profession's support to Hitler in 1933, the head of the two main German medical associations also outlined the economic problems of German doctors. Within a year non-Jewish doctors' incomes had risen by 25%, and within six years, Jewish doctors had been decertified and 95% had fled or were dead.

This is a fascinating and profoundly disquieting book, neatly exposing the tension between medicine as an ethical profession and as a business, and showing how readily the trumpeted human, and humane, values of medicine can be subverted by those of 'science' with perverted ends. It vividly illustrates Abraham Lincoln's remark that all men can deal with hardship but that the real test is how they deploy authority. Required reading for medical undergraduates? I think so. ■

John Galloway is at the Eastman Dental Hospital, 256 Gray's Inn Road, London WC1X 8LD, UK.

Phylogeny branches out

Phylogenetic Trees Made Easy: A How-to Manual for Molecular Biologists

by Barry G. Hall

Sinauer: 2001. 192 pp. \$24.95, £18.99 (pbk)

Yves Van de Peer

A little more than 10 years ago, when I was entering the field of phylogenetics myself, the construction of evolutionary trees was often a daunting task. In those, still early, days of personal computers, few programs for making trees were available, and it took about two or three days of fiddling with rulers, Chinese ink and typewriters to produce a publishable tree of less than 100 sequences. A lot has changed since then — currently, about 200 different software tools are available to construct and draw trees. Nevertheless, we had to wait a long time for a book like this. Whereas many other books or book chapters have focused on the theoretical aspects of certain methodologies or techniques, Barry Hall's is the first one, to my knowledge, that actually describes, step by step, how to build a tree.

Undoubtedly, there was a need for such a 'tutorial' book for non-expert users. Tree construction is a complicated business, and for those inexperienced in this matter the inference of phylogenies can be very frustrating, and the choice of programs and different options bewildering. Tree construction was, for a long time, mainly the business of biologists who studied the evolutionary relationship between organisms. In the era of comparative genomics, evolutionary analysis is becoming important in many new fields of

research. Phylogenies are used to deduce the relationships — originated by both speciation and duplication — between genes in large gene families across different species. Phylogenomics refers to the discipline that tries to improve functional predictions for uncharacterized genes by overlaying known functions of genes onto an evolutionary tree containing all homologues. Functions of uncharacterized genes are then predicted by their phylogenetic position relative to characterized genes. More and more, phylogenetic principles are used to interpret, and date, gene and genome duplications.

Basically, there are three main methodologies for inferring phylogenetic trees: maximum parsimony, pairwise distance methods and maximum likelihood. Maximum parsimony reconstructs the ancestral character states (nucleotides or amino acids) at the branching points of the tree, and 'chooses' the topology requiring the fewest number of changes to explain the sequences at the tips of the tree. Distance methods compute, for all pairs of sequences, their genetic distance — the fraction of sites that differ between two sequences, corrected for multiple substitutions. A tree is then inferred by considering the relationships among these distance values. Maximum-likelihood methods are statistical methods that try to find the tree topology that maximizes the probability of observing the data, again being the sequences at the tips of the tree. As with distance methods, likelihood methods are based on an explicit model of evolution; now, however, this model is used to compute the probability, or likelihood, of having an ancestral character state at a branching point in a given phylogeny.

Recently, as a fourth approach, bayesian inference has been applied to phylogenetic tree construction; this differs from maximum likelihood in that it depends on the 'posterior probability'. Posterior probabilities of trees are based on the joint probabilities of the tree, branch lengths and the model of substitution, and are approximated by sampling from the entire posterior-probability distribution (the so-called 'tree space', or collection of all possible tree topologies). Bayesian phylogenetic inference is indeed rapidly gaining popularity — partly because of the possibility of including prior information, such as the single origin of a group of sequences — and is obviously Hall's preferred method.

Hall explains how to build a tree from an alignment of sequences using these four different methodologies. In a stepwise manner, a publishable tree is created by making use of the most popular software packages available, such as PAUP* and TREE-PUZZLE.

The book unmistakably derives its strength from the clear examples (for which the data sets are available on the Internet) using software, and this is probably also its major disadvantage. Whereas PAUP* is available for PCs, and the Windows version runs

Barking up the right tree



Phaeographis inusta, found on the bark of deciduous trees and shrubs — from *Lichens of North America* by Irwin M. Brodo, Sylvia Duran Sharnoff and Stephen Sharnoff (Yale University Press, \$69.95, £50).

as a 'GUI' application, most options are, unlike the Macintosh version, command-line driven. For those who buy the book but have a PC, this will probably create some frustration, because the step-by-step procedures shown by displaying the menu options and different selections will be hard to execute in a command-line-driven environment. The book could have benefited from including the corresponding commands for the PC version. On the positive side, Hall discusses in detail aspects that may seem trivial, but often are not, such as the rooting and presentation of trees. Furthermore, a large part of the book is devoted to sequence alignment, and the author rightly emphasizes the importance of reliable alignment in tree building.

For many biologists, nothing is more exciting than seeing in a phylogenetic tree the sequences obtained after weeks or months of hard labour in the lab. This guidebook enables students and researchers anxious to start building trees to complete the job in a few minutes, and this will undoubtedly be highly appreciated. Unlike the pioneers, phylogenetic novices today have direct access to the appropriate computer hardware, software, and now even a tutorial manual that can initiate them into this important field of biology. ■

Yves Van de Peer is in the Department of Molecular and Plant Genetics, Flanders Interuniversity Institute of Biotechnology, Ghent University, B-9000 Ghent, Belgium.

Genetics of identity

Günter Theißen

Flowers are complex structures. They typically consist of four types of organ arranged in four whorls: sepals, petals, stamens and carpels. No wonder that their development is complicated. Nevertheless, there seem to be simple rules that underlie this process, as was realized when it became possible to analyse 'floral homeotic mutants' — plants with flowers that have seemingly normal floral organs in places where organs of another type are normally found.

In the model plant thale cress (*Arabidopsis thaliana*), homeotic mutants are categorized as A, B and C. Ideal class A mutants have carpels in the first whorl instead of sepals, and stamens in the second whorl instead of petals. Class B mutants have sepals rather than petals in the second whorl and carpels rather than stamens in the third; and class C mutants have petals instead of stamens in the third whorl and sepals instead of carpels in the fourth.

The existence of defined classes of mutants suggested simple combinatorial models, such as the 'ABC' model of flower development. This proposes three different floral homeotic functions to explain how the different floral organs adopt their unique identities during development. These functions are termed A, B and C, with A specifying sepals in the first floral whorl; A + B, petals in the second; B + C, stamens in the third; and C, carpels in the fourth. Because they had been identified by mutant analysis, it was clear that the homeotic functions are provided by sets of floral homeotic genes. In *Arabidopsis* there are at least two A-function genes, two B-function genes and one C-function gene, all of which encode

transcription factors — proteins that recognize specific DNA motifs and thereby regulate the expression of the genes that contain them.

Besides elegance, simplicity was certainly among the most attractive features of the ABC model, which found its way into modern textbooks and numerous reviews. However, it was soon realized that the model is incomplete. Expression of ABC genes throughout a plant does not transform leaves into floral organs. Thus the ABC functions, although necessary, are not sufficient to superimpose floral organ identity on a leaf development programme.

Another class of genes (*SEPALLATA*) is now known to be required and, together with the ABC genes, is sufficient for specification of petals, stamens and carpels. How do these *SEP* genes fit into the ABC model? Some have argued that they contribute to B and C functions, but they are more likely to have an additional function.

Indeed, despite these insights, I wonder whether floral homeotic functions are still a useful concept at all. Now that we know that there are at least as many floral homeotic functions as there are types of floral organ, the concept no longer provides a useful simplification. Wouldn't it suffice to refer to the four different states of floral organ identity on the one hand, and to the combinations of floral homeotic genes that specify these organ identities on the other? This seems to be a particularly useful way of thinking about the control of floral organ identity in the light of new evidence about how floral homeotic genes interact and exert their functions at the molecular level.

Proteins that are encoded by the floral

Flower development

The story of the outmoded ABC model shows that even the most successful concepts must be continually re-evaluated.

homeotic genes of *Arabidopsis* are now known to bind to DNA as multimeric complexes that contain B-function and SEP proteins, as well as either an A-function or a C-function protein. These are exactly the combinations of proteins that are sufficient to specify petal or stamen identity, respectively. This finding immediately suggests future research goals: to define the exact structures of the transcription-factor complexes inside the living plant cell; to identify the target genes for which transcription is regulated by the binding of the complexes; to explain the gene specificity of that binding; and to study how target-gene function brings about floral organ identity. Neither the concept of floral homeotic functions nor the ABC model will offer any help in achieving these exciting ambitions.

Was the ABC model an intellectual detour, or even an obstacle to scientific progress? Not at all. It provided a working hypothesis with numerous implications that could be experimentally tested, so it was an incredibly useful concept for a while. But scientific terms and concepts should be continually evaluated with respect to their heuristic value. Now that we have a direct understanding of the molecular mechanisms involved, there may no longer be a need for the abstract concept of floral homeotic functions to be squeezed between floral homeotic genes and floral organ identity. ABC functions may now obscure, rather than enlighten, our understanding of the linkage between molecular genetic events and floral phenotype.

The ABC model undoubtedly attracted many enthusiastic researchers to study flower development, and as such formed the basis for rapid progress in the field. So it could be that, ultimately, the success of the ABC model contributed to its abolition. ■

Günter Theißen is at the Institute for Genetics, Friedrich Schiller University, Philosophenweg 12, D-07743 Jena, Germany.

FURTHER READING

Egea Gutierrez-Cortines, M. & Davies, B. *Trends Plant Sci.* **5**, 471–476 (2000).

Goto, K., Kyoizuka, J. & Bowman, J. L. *Curr. Opin. Genet. Dev.* **11**, 449–456 (2001).

Jack, T. *Trends Plant Sci.* **6**, 310–316 (2001).

Theißen, G. & Saedler, H. *Nature* **409**, 469–471 (2001).

Floral combinations: the wildflower *Hepatica nobilis*



When one whale matters

Peter Kareiva

North Atlantic right whales once faced extinction and are still under threat today. But the population decline could be halted if the lives of just a few females were spared each year.

The northern right whale is one of the rarest mammals in the world. Hunted nearly to extinction in the nineteenth century, right whales have been protected for the past 65 years. Yet, after starting to recover, the North Atlantic population resumed its perilous decline around 1980. But the future for the right whale may not be as bleak as it appears: on page 537 of this issue, Fujiwara and Caswell¹ show that saving just two to three female whales each year could set this species on the road to recovery. Although it is widely known that right whales are in great peril, the conservation value of single animals has never before been so compellingly clear. Fujiwara and Caswell's analysis indicates that, if the recovery of this species is a high priority, every possible effort should be made to avert the death of adult females.

North Atlantic right whales (*Eubalaena glacialis*; Fig. 1) travel thousands of miles between feeding grounds off the coast of New England and calving grounds off the southeastern United States and even the Gulf of Mexico. Despite this wide dispersion, genetic analyses indicate that there is but one breeding population, which represents the descendants of possibly as few as three females². As is true for many conservation issues, field experiments are out of the question for investigating the decline of these whales. Instead, Fujiwara and Caswell¹ analysed a remarkable data set comprising over 10,000 individual sightings using the 'mark-recapture' technique. They performed numerical studies, manipulating birth and death rates, to explore which factors best account for the grim situation faced by these whales over the past two decades. Instead of simply lamenting some generic decline in survival, Fujiwara and Caswell pinpointed females, at a particular age, as the Achilles' heel for the survival of this species. Moreover, they show that random fluctuations in population size do not present a major source of risk.

Before this analysis, a prominent hypothesis cited to explain the decline was that North Atlantic right whales had been hunted to such low levels (fewer than 300 individuals) that females could not find mates in the vast Atlantic ocean. But this hypothesis is not supported by Fujiwara and Caswell's analysis. Instead, the problem appears to be that right whales, and particularly females, are not living long enough to get many chances



Figure 1 **Whale watching for conservation.** A sighting of a mother and calf right whale off the US coast.

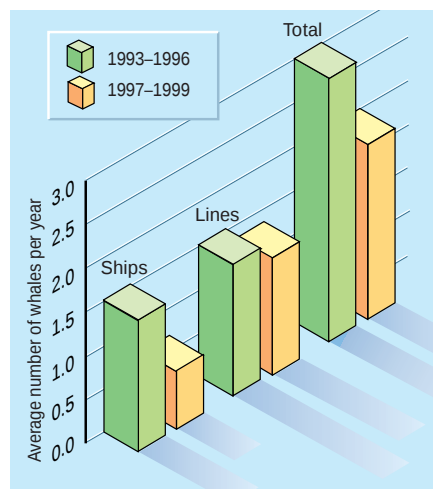


Figure 2 **Annual right whale injuries.** Rates of injury before (1993–1996) and after (1997–1999) the most recent US National Marine Fisheries Service (NMFS) regulations were imposed to reduce deaths of North Atlantic right whales. These records are taken from NMFS US Atlantic and Gulf of Mexico Marine Mammal Stock Assessments⁷ from 1999, 2000 and 2001. These are surely underestimates of injuries, yet the data draw attention to the fact that technical improvements in fisheries would go a long way to saving the two whales a year that would help this whale population to recover¹. Injuries are categorized as 'ship' if they involve deep gashes, crushed bones or other indications of impact. Injuries are categorized as 'lines' if there is evidence of having been entangled by fishing gear. Both injuries are likely to cause death.

at reproduction. Female right whales need nearly a decade to reach sexual maturity, and then produce only one calf every three to five years. In 1980, the average life expectancy of a female right whale was over 50 years, but by 1995 this had plummeted to less than 15. So, although females are not short of male partners, and can reproduce in any given year as well as they ever could, they simply do not

live long enough to give birth many times — down from over 5 chances at reproduction per female in 1980 to only 1.26 in 1995. The reasons for the decline in right whale life expectancy are unknown, but several factors, including human intervention and climate change, could be responsible. Mature females may be especially vulnerable because calving grounds — where reproductive

females spend a lot of time — tend to overlap with shipping lanes, and because the energy demands of motherhood exacerbate the ill-effects of any serious injury.

Because there are so few right whales left in the North Atlantic, the US National Marine Fisheries Service (NMFS) has already adopted a policy that emphasizes the value of each whale³. Consequently, although the NMFS has not attributed any special significance to mature females, they have interpreted the US Endangered Species Act of 1973 to mean that efforts must be taken to prevent any human-related whale deaths. Indeed, the NMFS initiated an extensive whale-tracking programme in 1996 to minimize such deaths by keeping ships away from sites of whale sightings⁴. But because more than 60% of North Atlantic right whales have scars from entanglement in fishing gear, such as lobster pots and sink gillnets (Fig. 2), environmental groups such as the Sierra Club in Massachusetts argue that simply alerting ships to the presence of whales does not provide adequate protection⁵. Moreover, the US Humane Society is currently suing the NMFS over its failure to address entanglement-related deaths, and has won an initial favourable ruling because the NMFS's current plan to minimize whale deaths is inadequately specified.

In short, Fujiwara and Caswell's study¹ provides both hope and despair. A population that was thought by many to be doomed because of terribly low numbers can probably be saved. But the fact that just a few human-induced deaths could tip the balance towards the population's demise exposes a familiar story: a government agency is charged with capitulating to a powerful industry, but claims to be doing all it can. Throughout the world, harvest of endangered species, whether accidental or deliberate, is permitted because the measures needed to prevent such deaths — such as closing down lobster fisheries — are seen as too draconian⁶. But according to environmental groups such as the Sierra Club, these extreme measures are not necessary⁵. They suggest instead that technical improvements in fishery operations may be sufficient, for example using lighter lines that break more easily when whales become entangled.

Population biology and demographic studies can predict how much better a population would fare if certain numbers of individuals were spared. Unfortunately, it is uncommon in conservation biology⁶ to get the kind of clarity provided by Fujiwara and Caswell's analysis. Conservation often means protecting individual animals from death caused by human activities. Fujiwara and Caswell have shown us the importance of highlighting what might at first glance seem like insignificant numbers of deaths. Their approach promises to be useful for everything from grizzly bears to spotted

owls to sockeye salmon — species for which similar mark-recapture data are available and which have in some places dwindled to such low numbers that saving individual lives could matter.

Peter Kareiva is at The Nature Conservancy, 217 Pine Street, Seattle, Washington 98101, USA.

e-mail: pkareiva@tnc.org

1. Fujiwara, M. & Caswell, H. *Nature* **414**, 537–541 (2001).
2. Hain, J. H. W. *USGS Technical Note*;

<http://biology.usgs.gov/s+t/SNT/noframe/mr183.htm>

3. Waring, G. T. *et al.* *NOAA Technical Memorandum NMFS-NE-153* (1999); <http://www.nefsc.nmfs.gov/psb/sar1999.pdf>
4. http://whale.wheelock.edu/whalenet-stuff/reportsRW_NE
5. McCaffrey, J. *The Massachusetts Sierran Online* **6** (2001); http://www.sierraclubmass.org/news/news0301/endangered_species.htm
6. Paine, R. T. *Report of the Recovery Science Review Panel* August 27–29, 2001; <http://www.nwfsc.noaa.gov/cbd/trt/rsrp.htm>
7. http://www.nmfs.noaa.gov/prot_res/PR2/Stock_Assessment_Program

Laser science

Physics at the attosecond frontier

Yaron Silberberg

Ultrashort laser pulses allow physicists and chemists to watch fast molecular motion as it happens. But many fundamental atomic processes are even faster and require the shortest pulses ever created.

As every photographer knows, a flash of light can stop the action. Just as a fast flash lamp can freeze the image of a bullet in mid-flight, so short laser pulses can be used to probe fast molecular motion. It is no surprise, then, that laser scientists have been pushing for ever-shorter pulses of light in order to follow ever-more rapid processes. The quest has taken us from the first sub-picosecond (1 picosecond is 10^{-12} s) pulses, more than a generation ago, to the development of femtosecond optics (1 femtosecond is 10^{-15} s). These time periods are hard to imagine, but a femtosecond is to a minute what a minute is to the age of the Universe.

Femtosecond pulses led to femtochemistry, the experimental study of fast chemical reactions and molecular dynamics. Even the fastest molecular vibrations appear com-

pletely still when probed with a pulse lasting a few femtoseconds. Now, we are entering the era of attosecond pulses (1 attosecond is 10^{-18} s). On page 509 of this issue Hentschel *et al.* describe¹ the generation and use of pulses lasting 650 attoseconds, in what might be the dawn of attophysics — the study of dynamics on timescales fast enough to follow electronic motion within atoms.

The road from picosecond to femtosecond light pulses has seen laser technology evolve towards lasers that emit light with greater 'spectral' width — that is, covering a wider range of wavelengths. A short pulse results when all the spectral components in the light beam interfere in a way that adds up to a single burst of light. The duration of this pulse is inversely proportional to the spectral width — so a wider spectral

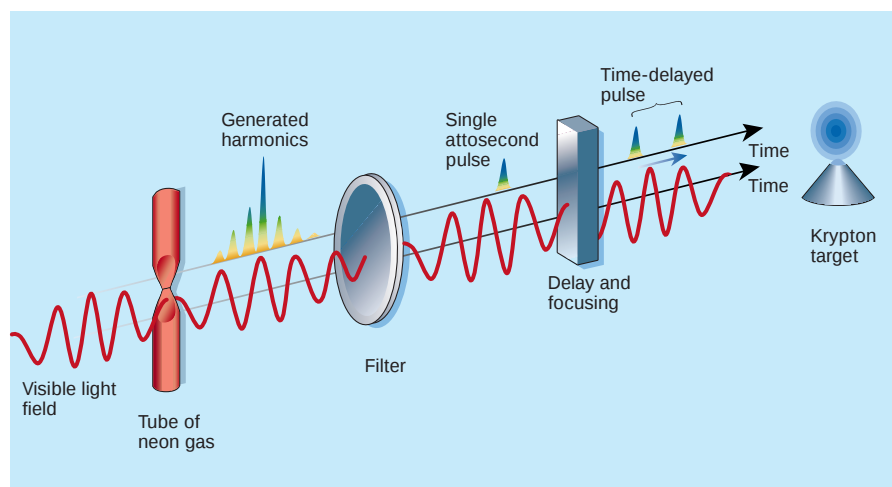


Figure 1 Generating and using attosecond laser pulses. In their experiment, Hentschel *et al.*¹ shine an ultrashort visible light pulse on a gas of neon atoms to produce higher frequency 'harmonic' radiation (rainbow-coloured pulses) at ultraviolet and X-ray wavelengths. The visible and harmonic pulses pass together through a zirconium filter, which reduces the train of harmonic pulses to a single attosecond pulse (1 attosecond is 10^{-18} s). Both the attosecond pulse and the optical beam are focused onto the krypton target where they accomplish the very first attosecond measurement. The authors monitor the attosecond dynamics of photoelectrons emitted by the krypton gas by varying the delay between the attosecond and visible pulses.

span means a shorter pulse. This approach reached its natural limit when the spectral span of the laser became a significant portion of the visible spectrum. This is because the pulse cannot be shorter than the period of the optical wave, which is a few femtoseconds in the visible range. So it was obvious that a different approach would be needed to break into the attosecond range.

The latest advances are based on research into 'high-order harmonics' of femtosecond laser pulses. Nearly a decade ago^{2,3} physicists used intense femtosecond pulses to ionize a rare gas (such as neon) and found that new electromagnetic waves were generated at much higher frequencies (and shorter wavelengths) alongside the original optical pulse (at visible and short infrared wavelengths). It is fairly straightforward to produce these high-order harmonics, whose wavelengths extend well into the ultraviolet and X-ray range, but measuring and controlling them to produce ultrashort pulses is much harder.

Precise analysis of the way the gas atoms interact with the laser field requires careful application of quantum mechanics, but a simple model developed in 1993 by Paul Corkum⁴ (one of the authors of the Hentschel *et al.* paper) provides some insight. In this model, the original laser pulse tears an electron away from an atom, and the freed electron moves in response to the laser field, being accelerated and decelerated as the electromagnetic wave oscillates. The new harmonics are generated when the electron collides with the ion it left behind. This radiation is termed harmonic because its frequencies are multiples of the original laser frequency.

Corkum argued that electrons that are ejected precisely at the peaks or the crests of the optical pulse are much more likely to radiate. So the radiation is produced in very short bursts, which occupy just a fraction of the optical cycle. These short bursts explain the much wider spectral span of the high-order harmonics. The bursts are estimated to last some 100 attoseconds, and are expected to appear twice in each optical cycle of the original pulse. Last year, two experiments confirmed that the duration of these pulses is in the attosecond range^{5,6}.

Attosecond pulses produced in this way appear as a train of pulses separated from each other by half an optical wavelength. However, most experiments require isolated pulses. The key to isolating a single attosecond pulse is to start with a very short pulse of just a few optical periods. Such a pulse produces only a few attosecond bursts, and it is possible to select just one by appropriate filtering (Fig. 1). So Hentschel *et al.*¹ start with a 7-femtosecond optical pulse and estimate that, after filtering, more than 90% of the energy of the new radiation they produce is contained in a single 650-attosecond pulse.

But that's not all. After generating an

isolated attosecond pulse, Hentschel *et al.* shine it, together with the optical pulse that generated it, onto krypton gas, to perform the first attosecond measurements of an electronic process. While varying the delay between the two pulses with attosecond accuracy they monitor the energy spectrum of any photo-electrons ejected from the gas. (Changing the difference between the paths taken by the two pulses by 0.3 micrometres changes the delay between the two pulses by 1 femtosecond, so it is easy to control the pulse timing in this way.) With this measurement Hentschel *et al.* achieve three important things: they characterize the attosecond pulse; they measure the detailed structure of the optical field; and they perform the first sub-femtosecond measurement of electron dynamics associated with photo-ionization.

What sort of experiments are possible with attosecond pulses? Should we expect an explosion of new results? At these timescales, chemistry is essentially frozen in time, so the only dynamics to be studied are those of electrons, as they are much lighter and faster than nuclei. Most femtochemistry measurements are based on the pump-probe method: one pulse sets the system into motion, and a second one probes it after a controlled delay. But this approach is not possible with attosecond pulses: they are just too weak. Moreover, manipulating

such pump and probe pulses is harder to do because their wavelength is in the X-ray range, for which traditional optical tools (such as lenses and beamsplitters) do not exist.

The work of Hentschel *et al.* hints at the likely direction that attophysics will take in the near future. They achieve attosecond measurements by carefully timing the delay between an attosecond pulse and an optical pulse. Yet the experiment also demonstrates the difficulties associated with this approach, because the information in the attosecond pulse, the optical field and the physics of the interaction are all interwoven. The interpretation of such experiments will necessarily be complex, and any physical information will have to be derived indirectly. We have now entered the attosecond world, but we will need a better guidebook to help us find our way around it. ■

Yaron Silberberg is in the Department of Physics of Complex Systems, Weizmann Institute of Science, 76100 Rehovot, Israel.

e-mail: yaron.silberberg@weizmann.ac.il

1. Hentschel, M. *et al.* *Nature* **414**, 509–513 (2001).
2. L'Huillier, A. & Balbou, P. *Phys. Rev. Lett.* **70**, 774–777 (1993).
3. Macklin, J. J., Kmetec, J. D. & Gordon, C. L. III *Phys. Rev. Lett.* **70**, 766–769 (1993).
4. Corkum, P. B. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
5. Papadogiannis, N. A. *et al.* *Phys. Rev. Lett.* **83**, 4289–4292 (1999).
6. Paul, P. *et al.* *Science* **292**, 1689–1692 (2001).

Oceanography

Sea snow microcosms

Farooq Azam and Richard A. Long

Marine bacteria can respond to organic particles in sea water, creating hotspots of bacterial growth and carbon cycling. This microscale behaviour should be included in models of the oceanic carbon cycle.

Oceanographers exploring the ocean's carbon cycle, and its role in climate change, do so at a grand scale. Satellite measurements of ocean colour are used to infer carbon dioxide (CO₂) assimilation into organic matter by photosynthetic organisms — a process known as carbon fixation. Instruments on ships and buoys assess how fast carbon is being fixed, what fraction is respired back by marine organisms to CO₂, and what fraction of organic matter sinks into the deep ocean, isolating its carbon from air-sea exchange for millennia. These studies paint a broad-brush picture of carbon cycling in the oceans, but they cannot predict future changes — in response to increases in atmospheric CO₂, for example.

Without looking at the ocean at the finer scale we cannot hope to understand or predict these changes^{1,2}. It has long been recognized that bacteria are responsible for consuming much of the organic matter in marine systems. But the current climate-

change models ignore how bacteria in the ocean respond to the patchy distribution of organic matter. This may well be a serious omission. A simulation, reported in *Limnology and Oceanography* by Kiørboe and Jackson³, greatly reinforces our understanding of how marine bacteria interact with the patchy distribution of organic matter in the ocean. The results indicate that the response of bacteria to sinking organic particles has a big influence on the oceanic carbon cycle.

Kiørboe and Jackson³ model the interactions of bacteria with marine snow (Fig. 1, overleaf). A cubic metre of sea water contains thousands of these rapidly sinking flakes⁴. They are the main vehicle for transporting organic matter, and its associated carbon, from the upper ocean⁴ (where it readily exchanges with the atmosphere) to the ocean floor (where it eventually settles in sediments). The fate of marine snow, and the depth at which its carbon is converted back to CO₂ by bacteria, is critical for under-

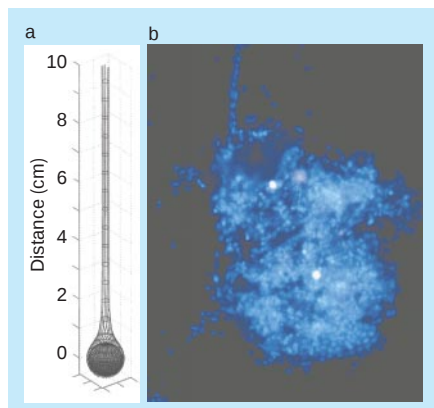


Figure 1 Marine snow. a, In their model Kiørboe and Jackson³ consider the formation of a plume of dissolved organic matter streaming behind a sinking particle of marine snow. The plume is formed by the enzymatic dissolution of the snow particle by colonizing bacteria. b, Marine snow collected by a diver in Monterey Bay, California. The marine snow is intensively colonized by bacteria ($\sim 10^9$ bacteria per millilitre) that have been stained blue with the fluorochrome DAPI. A fragment of snow was photographed in an epifluorescence microscope⁹. The photograph shows a $37\ \mu\text{m} \times 27\ \mu\text{m}$ segment from a marine snow aggregate.

standing whether the ocean will act as a source or sink of carbon in future climate change scenarios. It turns out that the answer greatly depends on whether or not bacteria can sense and respond to the presence of marine snow.

Bacteria that consume organic matter are abundant in the upper ocean, achieving densities of around 10^6 per ml, but they can only absorb dissolved molecules. So they must first convert marine snow into dissolved organic matter (DOM). There is increasing evidence that bacteria in sea water can sense chemical gradients and are attracted towards nutrient molecules⁵⁻⁷; many can swim hundreds of micrometres per second — equivalent to hundreds of body lengths per second⁸. Bacteria attracted to the organic particles intensively colonize marine snow (reaching densities of 10^9 per ml) and produce copious extracellular enzymes, whose activities release and break down components such as proteins and polysaccharides from the snow, bringing small nutrient molecules into solution⁹. This converts the sinking organic matter into non-sinking DOM and keeps it in the upper ocean. But the colonizing bacteria produce DOM much faster than they can use it, so the sinking marine snow leaves a rich plume of DOM in its wake⁹. Kiørboe and Jackson³ predict that free-living bacteria can take advantage of this plume and use it to fuel a population explosion. For example, bacteria in bulk sea water will double in numbers every 1–10 days, whereas bacteria living in the plume can double in just a few hours to a day.

So Kiørboe and Jackson's model³ predicts that bacteria in the plume grow much faster than those in bulk sea water. But perhaps equally important is the increased efficiency of bacterial growth in the plume, defined as the fraction of DOM carbon that is assimilated as bacterial biomass (weight). In general, bacteria growing rapidly, as in the plume³, have higher efficiency than those growing slowly in bulk sea water¹⁰. So, bacterial use of DOM within nutrient-rich plumes could substantially lower total CO_2 respiration by the marine ecosystem by fixing a higher fraction of carbon as biomass.

According to Kiørboe and Jackson, nutrient-rich plumes produced from marine snow could support most heterotrophic bacterial biomass production, despite occupying a small fraction of total sea water. These plumes would attract bacteria-eating protozoa and larger animals that feed on protozoa. Thus, the plumes might be the focus of complex ecosystems (Fig. 2), with implications for food web and fisheries models. Marine snow is also enriched in iron and its breakdown and release of iron into the plume might create iron-rich patches in the oceans. This aspect of iron cycling must be considered in carbon-cycle models that assume that the growth of plankton is limited by iron supply.

It is easy to see the benefits gained by free-living bacteria that swim into the plume, but the behaviour of the bacteria that colonize the snow particles seems paradoxical. They expend energy swimming towards, colonizing and synthesizing enzymes to dissolve marine snow, but the released DOM enables free-living bacteria, not the colonizing bacteria, to achieve explosive growth. Growth rates of colonizers themselves are quite modest⁹. Furthermore, the colonizers

face difficulties not faced by the plume-seekers: sinking to depths to which they are not adapted (low temperature, high pressure) and being eaten by animals. Perhaps the paradox can be explained by assuming that snow-colonizers release their progeny into the plume (Fig. 2), where they can enjoy explosive growth, remain in the upper ocean, and escape predation by animals that consume marine snow.

A logical next step in building on the model of Kiørboe and Jackson would be to introduce the complexity of the species composition of the bacteria — both the colonizers and those in plumes. Different colonizing bacteria can produce different enzyme activities¹¹, and this would be reflected in the composition and concentration of DOM in the plume. Furthermore, the uptake capabilities of bacteria in the plumes will determine which components of the DOM plume they can absorb and how efficiently. The composition of DOM may, in turn, make the plume more attractive for certain species. It would be worth investigating whether human pathogens and sewage-derived enteric bacteria might also sense and swim into snow plumes and experience explosive growth.

Kiørboe and Jackson's paper³ deals with only one type of organic matter, but sea water is replete with even finer particles¹ (Fig. 2). At these small scales, the heterogeneity in the distribution of organic matter, and bacterial responses to it, are likely to fundamentally influence the structure, function and stability of marine ecosystems. Commonly used continuum models of marine ecosystems cannot capture this heterogeneity, giving biased interpretations of the underlying dynamics¹². Kiørboe and Jackson's paper reminds us that much remains to be discovered before

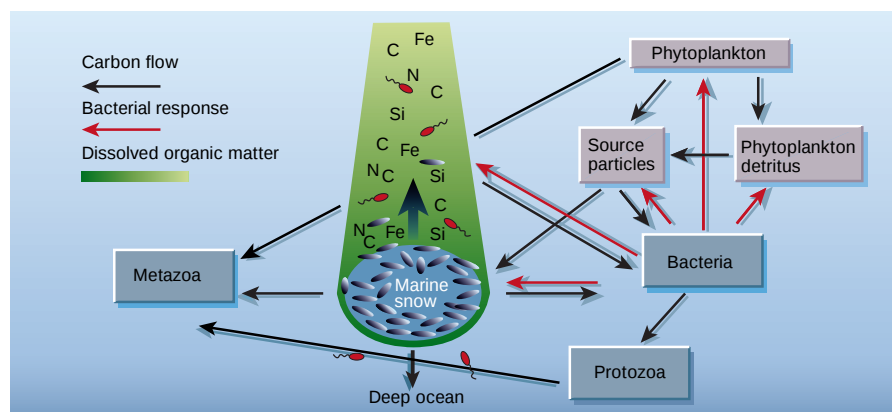


Figure 2 Carbon fluxes in the ocean involving marine snow. The marine snow (aggregated phytoplankton and dead material) can sink into the deep ocean, thereby removing it from the upper ocean, or it can become involved in nutrient cycling. The first step is colonization by bacteria, which produce enzymes that turn the marine snow into dissolved organic matter (DOM). The colonizing bacteria produce DOM faster than they can use it, so the sinking snow releases a plume of material containing carbon (C), nitrogen (N), silicon (Si) and iron (Fe). Other free-living bacteria (red) are attracted to the plume and grow rapidly. Colonizers may also release their progeny (blue) into the plume. High concentrations of bacteria attract protozoa, which in turn attract larger animals (metazoa). The marine snow and the plume may thus become the focus of a complex food web.



100 YEARS AGO

In "Monkeys; their Affinities and Distribution," by Dr. A. R. Wallace, the author gives as one of the characters in which man *differs* from all the monkey tribe — "*the perfect freedom of the hands from all part in locomotion.*" My object in writing this is to point out the peculiar way in which the majority of people move their arms and hands when walking or running. One may safely say that everybody, adults and children, at one time or another exercise this movement. The natural way in which children run is to "paddle" with the arms and hands, though trained runners do not do so. Now is it not possible that this muscular movement of the fore-limbs in opposite directions in the act of locomotion is a survival of the four-legged mode of progression of man's remote ancestors?... I believe that this theory has been thought of before, but I am unable to find any trace of it in the books I have consulted. I should be very grateful if any of your readers would enlighten me on the subject. *Basil W. Martin*
From *Nature* 28 November 1901.

50 YEARS AGO

The story of the rabbit in Australia is so well known, so familiar an instance of the spread of an innocent introduction to plague dimensions, that fresh turns in its course must be spectacular to become news... The recognition that the virus disease, myxomatosis, might be an agent in reducing rabbit numbers directly invoked scientific enquiry and although early investigations... established that the disease could work, its effectiveness was small because of its very limited powers of spreading naturally through rabbit populations and groups... But as [this report] was prepared, a spectacular development occurred in the development of myxomatosis epidemics. It was known that the disease could be spread from sick to healthy rabbits by certain mosquitoes. As mentioned above, the first attempts to establish the disease in experimental sites were unsuccessful, but by continuing the introduction into warren colonies in the Murray River valley into the summer months of 1950–51, conditions favourable to success were encountered and mosquitoes rapidly spread the disease, so that by February 1951 the disease was reported widely in the Murray River system and its incidence had extended northwards along the Murrumbidgee, the Lachlan and the Darling Rivers, almost up to the Queensland border. From *Nature* 1 December 1951.

we fully understand the mechanisms behind microbial control of carbon cycling. Given our current state of ignorance, any large-scale manipulation of the oceans, such as proposals to increase phytoplankton growth by iron fertilization, could have catastrophic consequences^{13,14}. There is an urgent need for oceanographers to embark on a bold exploration of the oceans — this time at the millimetre scale.

Farooq Azam and Richard A. Long are in the Marine Biology Research Division, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92092-0202, USA.

e-mail: fazam@ucsd.edu

1. Azam, F. *Science* **280**, 694–696 (1998).
2. Bidle, K. D. & Azam, F. *Nature* **397**, 508–512 (1999).

3. Kiorboe, T. & Jackson, G. A. *Limnol. Oceanogr.* **46**, 1309–1318 (2001).
4. Alldredge, A. L. & Gotschalk, C. C. *Cont. Shelf Res.* **10**, 41–58 (1990).
5. Mitchell, J. G., Pearson, L. & Dillon, S. *Appl. Environ. Microbiol.* **62**, 3716–3721 (1996).
6. Blackburn, N., Fenchel, T. & Mitchell, J. *Science* **282**, 2254–2256 (1998).
7. Grossart, H.-P., Riemann, L. & Azam, F. *Aquat. Microb. Ecol.* **25**, 247–258 (2001).
8. Mitchell, J. G., Pearson, L., Dillon, S. & Kantalis, K. *Appl. Environ. Microbiol.* **61**, 4436–4440 (1995).
9. Smith, D. C., Simon, M., Alldredge, A. L. & Azam, F. *Nature* **359**, 139–142 (1992).
10. del Giorgio, P. A. & Cole, J. J. *Annu. Rev. Ecol. Syst.* **29**, 503–541 (1998).
11. Martinez, J. *et al.* *Aquat. Microb. Ecol.* **10**, 223–230 (1996).
12. Young, W. R., Roberts, A. J. & Stuhne, G. *Nature* **412**, 328–331 (2001).
13. Fuhrman, J. A. & Capone, D. G. *Limnol. Oceanogr.* **36**, 1951–1959 (1991).
14. Chisholm, S. W., Falkowski, P. G. & Cullen, J. J. *Science* **294**, 309–310 (2001).

Cell cycle

Six steps to destruction

James E. Ferrell Jr

Cell division relies on the properly timed activation and destruction of certain regulatory proteins. New work shows that many rounds of phosphorylation can help to establish the timing of protein destruction.

The process by which cells multiply consists of a strictly ordered series of events. A cell copies its genetic material, grows, and segregates the duplicated DNA into two new cells. Then, after a rest period, the process begins again. If cells are to progress smoothly from one phase of this 'cell cycle' to the next, key regulatory molecules must be turned on and off at just the right times. On page 514 of this issue, Nash and colleagues¹ provide new insight into the molecular controls that allow budding yeast cells to move from the G1 phase (the rest period) to the S phase (when DNA is copied). Their work builds on the fact that the Sic1 protein needs to be enzymatically modified with phosphate groups (phosphorylated) over and over again before cells can progress

from G1 into the S phase. Sic1 provides a splendid example of how cells can use cumulative, multistep modifications to produce a strict, switch-like transition.

The job of the Sic1 protein is to inhibit the protein complex that drives budding yeast cells from the G1 phase to the S phase^{2–4}. This protein complex comprises an enzyme from a family known as the cyclin-dependent kinases (this enzyme is Cdc28 in yeast, and Cdk1 in mammals) and its regulatory subunit, a cyclin protein (from the Clb family in yeast). During the G1 phase, cells have high levels of Sic1; this ensures that Cdc28–Clb complexes are inactive, so DNA replication is suppressed. At the end of the G1 phase, Sic1 is degraded, allowing Cdc28–Clb to drive cells into S phase. In yeast strains that lack

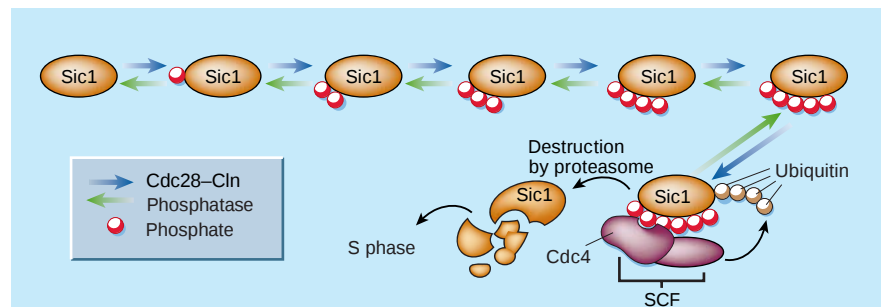


Figure 1 How cells get from the G1 phase to the S phase in their division cycle. The shift from G1 to S phase is blocked by the protein Sic1, which inhibits a key enzymatic complex. Nash *et al.*¹ have discovered that Sic1 must be phosphorylated at least six times by the Cdc28–Cln complex before it can bind to the Cdc4 protein, be tagged with ubiquitin groups, and destroyed. The multistep phosphorylation of Sic1 may function as a sort of primordial timing mechanism, ensuring that Sic1 destruction is precisely timed so that the S phase does not happen too early.

Sic1, DNA replication occurs too early; in strains with forms of this protein that are hard to destroy, DNA replication is delayed. So the timing of Sic1 degradation helps to establish the timing of the G1/S transition.

Before Sic1 can be degraded, however, it must first be phosphorylated. The enzymes that do this comprise Cdc28 (again) and a Cln protein (another cyclin), and are activated in the G1 phase^{5,6}. Phosphorylation allows Sic1 to bind to the protein Cdc4, which is a component of yet another protein complex, called SCF^{7,8}. SCF in turn ensures that Sic1 becomes modified with a small peptide, ubiquitin, and targeted to the cellular protein-degrading machinery, the proteasome. The timing of Sic1 destruction therefore depends on the timing of its phosphorylation.

Sic1 has nine different sites that fit the 'consensus' for phosphorylation by a cyclin-dependent kinase (the consensus is the amino-acid sequence serine-proline or threonine-proline), and most, if not all, of the sites appear to become phosphorylated *in vivo*⁶. This raises the question of why Sic1 has so many phosphorylation sites and is phosphorylated so many times.

Nash and colleagues¹ approached this question by mutating different numbers of phosphorylation sites to find out how many are needed for proper destruction of Sic1. They found that Sic1 with two, three or even five sites could not bind Cdc4 *in vitro*, but Sic1 with six, seven or nine sites could. Likewise, Sic1 with just two, three or five phosphorylation sites prevented yeast strains from multiplying — presumably because they arrested in the G1 phase — but strains containing Sic1 with six, seven or nine sites divided normally.

Why isn't one phosphorylation (or even five) sufficient? One possibility is that Cdc4 has six phosphate-binding pockets. But quantitative binding studies and mutational analysis indicate that it has only one¹. An alternative is suggested by the consensus amino-acid sequence within phosphorylated peptides that allows them to bind optimally to Cdc4 (ref. 1). The mammalian cyclin E protein, another SCF substrate, has a single phosphorylation site that targets it for degradation; the sequence around this site fits the Cdc4-binding consensus and binds Cdc4 with high affinity. But it turns out that none of the phosphorylation sites in Sic1 matches this consensus very well. This implies that Sic1 needs to be phosphorylated many times because none of its phosphorylation sites is all that good at binding Cdc4.

But is it important that Sic1's destruction is driven by the combined effects of six phosphorylation sites? Or would a single high-affinity site work just as well? To find out, Nash *et al.* replaced seven of Sic1's nine phosphorylation sites with a single high-

affinity site derived from cyclin E. The resulting protein inhibited Cdc28–Cln *in vitro*, but failed to restrain DNA replication properly *in vivo*. Apparently it really does matter how Sic1 comes to be destroyed.

The fact that Sic1 does not appreciably bind Cdc4 and begin to be destroyed until the sixth phosphate is added has several kinetic implications. It means that there should be a discrete lag time between the activation of Cdc28–Clns and the destruction of Sic1 (Fig. 1). Those ineffectual first five phosphorylations need to be accomplished before the critical sixth one can occur. So the time required for the first five phosphorylations acts as a temporal threshold for Cdc28–Cln activation. In one sense, the first five phosphorylations accomplish nothing — but sometimes doing nothing is of the utmost importance.

The requirement for six phosphates also means that the destruction of Sic1 could be a highly nonlinear function of the amount of Cdc28–Cln activity in a cell. If so, the regulation of the G1/S transition would have a noise filter built into it — low levels of Cdc28–Cln activity would result in very little Sic1 being destroyed. Sic1 would then respond to higher levels of Cdc28–Cln activity in a decisive, switch-like fashion; doubling Cdc28–Cln activity could increase the destruction of Sic1 by a factor of 2⁶ (64-fold)⁹.

The six phosphorylations may also help to ensure the specificity of Sic1 destruction. Suppose that some kinase other than Cdc28–Cln can inappropriately phosphorylate Sic1 at a low frequency, ϵ . The frequency of inappropriate destruction of Sic1 would then be as low as ϵ^6 . This property has been

termed 'kinetic proof-reading'¹⁰ (P. Swain and E. Siggia, personal communication), and is another potential advantage of the six-phosphate mechanism.

Not all substrates of the SCF complex need multiple phosphorylations for destruction; cyclin E is a good example of this. So the timing and abruptness of destruction seems to be regulated by the properties both of the enzymes that catalyse phosphorylation and destruction, and of their substrates. SCF substrates apparently have a primordial timing mechanism built into them, on top of which other controls can be built.

Studies of budding yeast have a long history of yielding valuable insights into cell biology. Nash and co-workers¹ have shown that such studies can also shed light on the biochemical logic of cell regulatory systems — with a lot of hard work, a bit of quantitative reasoning, and the ability to count to six. ■

James E. Ferrell Jr is in the Departments of Molecular Pharmacology and Biochemistry, Stanford University School of Medicine, Stanford, California 94305-5174, USA.
e-mail: james.ferrell@stanford.edu

1. Nash, P. *et al.* *Nature* **414**, 514–521 (2001).
2. Schwob, E., Bohm, T., Mendenhall, M. D. & Nasmyth, K. *Cell* **79**, 233–244 (1994).
3. Mendenhall, M. D., al-Jumaily, W. & Nugroho, T. T. *Prog. Cell Cycle Res.* **1**, 173–185 (1995).
4. Nugroho, T. T. & Mendenhall, M. D. *Mol. Cell. Biol.* **14**, 3320–3328 (1994).
5. Schneider, B. L., Yang, Q. H. & Futcher, A. B. *Science* **272**, 560–562 (1996).
6. Verma, R. *et al.* *Science* **278**, 455–460 (1997).
7. Feldman, R. M., Correll, C. C., Kaplan, K. B. & Deshaies, R. J. *Cell* **91**, 221–230 (1997).
8. Skowrya, D., Craig, K. L., Tyers, M., Elledge, S. J. & Harper, J. W. *Cell* **91**, 209–219 (1997).
9. Ferrell, J. E. Jr *Trends Biochem. Sci.* **21**, 460–466 (1996).
10. Hopfield, J. J. *Proc. Natl Acad. Sci. USA* **71**, 4135–4139 (1974).

High-energy astrophysics

A new spin on black-hole masses

Charles Bailyn

The extreme environment surrounding a black hole provides an ideal test bed for the predictions of general relativity. New observations of a spinning black hole push current theories to their limits.

One of the most enigmatic X-ray sources in our Galaxy is the prosaically named GRS1915 + 105, a type of binary star system known as a microquasar. It is of unique importance in high-energy astrophysics because it is the only microquasar that has remained active for more than a few months. Since its discovery¹ in 1994, there have been hundreds of papers describing and interpreting its X-ray, infrared and radio emission. Only the optical astronomers have been left out of the fun because a thick layer of galactic dust blocks optical light emitted by the source. The lack of optical data is unfortunate, because optical obser-

ventions (at visible and short infrared wavelengths) are usually required to provide crucial information about binary star systems, such as the orbital period and masses of the companion stars. On page 522 of this issue, Greiner *et al.*² at last report infrared observations of GRS1915 + 105 that enable them to determine these basic parameters. The results confirm the nature of the two companions — one is a mature Sun-like star, the other a spinning black hole — but the mass of the black hole gives an unexpected twist to the story.

Much of the excitement generated by GRS1915 + 105 is due to its 'superluminal'



Figure 1 Artist's impression of superluminal jets emanating from a 'microquasar'. Similar jets are seen emanating from quasars, also known as active galactic nuclei, containing supermassive black holes. So similar objects found in our Galaxy, like the one studied by Greiner *et al.*² that contains a spinning black hole, have been called microquasars.

jets. In radio observations, narrow jets of material can be seen squirting out of the immediate vicinity of the black hole, and travelling across the sky at speeds apparently faster than the speed of light³. The apparent speed is a well-known optical illusion caused by special relativity, but the true velocity of the jet material is calculated to be greater than 90% of the speed of light. For decades, superluminal jets have been seen emanating from the vicinity of quasars, now known to be supermassive black holes at the centres of galaxies. So objects such as GRS1915 + 105 have been dubbed microquasars (Fig. 1).

Microquasars are natural laboratories for testing the predictions of Einstein's general theory of relativity under conditions in which there are extremely strong gravitational fields. In general relativity, Newton's idea that gravity is a force is replaced by the concept of a four-dimensional space-time geometry, in which the three dimensions of space and one of time are represented. The presence of gravitational fields generates 'curvature' in this space-time geometry, and this curvature changes the trajectories of objects embedded within it. General relativity has already proved itself in situations that deviate only slightly from newtonian physics, such as the Solar System, and in binary pulsars, where gravitational fields are fairly modest.

The extreme gravitational fields generated near the event horizon of a black hole, especially one that is spinning rapidly and is close to a visible companion, make microquasars prime candidates for studying relativistic effects. In particular, the twisting of space-time caused by the spin of the black hole may be responsible for narrowly

focusing the superluminal jets. In this model, a mechanism for accelerating the jets to near-light speed requires the black hole to rotate at close to the maximum rate allowed by general relativity. So microquasars are likely to have higher spin rates than similar systems without superluminal jets⁴, and the relativistic effects associated with spinning black holes may be more conspicuous.

A striking effect that might be caused by the spin of the black hole is the so-called quasi-periodic oscillations or QPOs, seen in the emission of GRS1915 + 105 and many other X-ray sources. In an X-ray binary system, the black hole accretes material from the companion star, generating X rays as the material swirling around the black hole heats up. X-ray observations of these binary systems reveal intensity variations of a characteristic frequency, sometimes as rapid as hundreds of hertz, but they are not strictly periodic — hence the name 'quasi-periodic' oscillations.

There are many different QPOs, created by various physical processes, but a particular QPO pattern seen in microquasars has been associated with the spin frequency of the black hole⁴. A number of theories try to explain the relationship between the QPO frequency and the spin frequency of the black hole, most of which suggest that for maximally rotating black holes in microquasars the QPO frequency should be inversely proportional to the mass of the black hole. So far, astronomers have been able to obtain both the QPO frequency and the mass of a black hole for only one microquasar⁵, known as GRO J1655-40. On the basis of their theories, and the data from GRO J1655-40, many astronomers expect-

ed that the black hole in GRS1915 + 105 should be four times the mass of the black hole⁶ in GRO J1655-40 (which is 5–7 solar masses) because its QPOs are four times slower⁷.

Until now, the mass of the black hole in GRS1915 + 105, in principle a straightforward measurement, remained unknown. Repeated measurements of spectral lines from the companion star are used to determine both the orbital period and a projected velocity for the companion (calculating the true velocity also requires knowing the orbital inclination). By applying Kepler's laws to the velocity, orbital period and ratio of masses in the system (obtainable by other means) one gets a value for the mass of the unseen component of the binary system, in this case the black hole. Generally, such measurements are made in optical light, where most of the energy from the Sun-like companion stars is emitted, but this has not been possible for GRS1915 + 105.

Using one of the four newly constructed 8-metre telescopes of the European Southern Observatory, Greiner *et al.*² have succeeded in identifying spectral features of the companion star in the infrared emission of GRS1915 + 105. By monitoring these features over a period of months, they determined the orbital period and projected velocity of the companion, while observations of the jets provided the orbital inclination. Now they can calculate the mass of the black hole in the GRS1915 + 105 system.

The result, 14 ± 4 solar masses, is not what astronomers had predicted on the basis of the QPO frequency. As Greiner *et al.* show, this mass value, just over twice the mass of GRO J1655-40, is inconsistent with any of the current models of QPO production in microquasars. It may be that we cannot assume that microquasars that emit jets are spinning close to the maximum allowed rate, a result that would require a rethink of many jet-formation theories. Or it may be that the QPO frequency is not related to the spin frequency of the black hole, or is related in some as-yet unexpected way. What is almost certain is that this sudden detour in the study of relativistic effects in microquasars will generate several hundred more papers in the years to come.

Charles Bailyn is in the Department of Astronomy, Yale University, PO Box 208101, New Haven, Connecticut 06520-8101, USA.
e-mail: bailyn@astro.yale.edu

1. Castro-Tirado, A. J. *et al. Astrophys. J.* **92**, 469–472 (1994).
2. Greiner, J., Cuby, J. G. & McCaughrean, M. *J. Nature* **414**, 522–525 (2001).
3. Mirabel, I. F. & Rodríguez, L. F. *Nature* **371**, 46–48 (1994).
4. Cui, W., Zhang, S. N. & Chen W. *Astrophys. J.* **492**, L53–L55 (1998).
5. Remillard, R. A., Morgan, E. H., McClintock, J. E., Bailyn, C. D. & Orosz, J. A. *Astrophys. J.* **522**, 397–412 (1999).
6. Greene, J., Bailyn, C. D. & Orosz, J. A. *Astrophys. J.* **554**, 1290–1295 (2001).
7. Morgan, E. H., Remillard, R. A. & Greiner, J. *Astrophys. J.* **482**, 993–1010 (1997).

Ion channels

Accessory to kidney disease

Malcolm Hunter

The protein that is mutated in a human disorder of the kidney and ear turns out to be an accessory subunit for a chloride ion channel. The discovery explains the symptoms of the disease.

Kidneys are filters: they remove waste substances from the bloodstream, producing urine. In so doing, they also regulate various qualities of the blood, such as its salt level. When kidneys malfunction the results can be debilitating, as occurs in Bartter's syndrome, a salt-wasting disorder that affects humans. Bartter's syndrome was first described nearly 40 years ago¹, and several subtypes, caused by different genetic defects, have since been identified. One of these, Bartter's syndrome type IV, is also associated with deafness². The type IV syndrome is quite rare, affecting just one in a million live births. Patients are generally identified at birth or in early childhood, with the main symptoms — apart from deafness — being increased urine flow, a great thirst (to cope with the loss of fluid in urine), and a failure to thrive. The gene mutated in the type IV syndrome was identified earlier this month, and encodes a small membrane protein christened barttin³. Estévez and colleagues⁴ have now found out what this protein does, as they describe on page 558 of this issue.

Each human kidney is made up of roughly half a million hollow tubules called nephrons. Blood plasma is filtered into the

nephrons and is then modified by the selective absorption or secretion of various ions or molecules by the cells that make up the walls of the tubules. The fluid that remains in the tubules forms the urine. The function of the tubule cells varies along the length of a nephron. For example, one part of the nephron — the thick ascending limb — actively absorbs sodium and chloride ions, creating an osmotic gradient that allows water to be reabsorbed into the body if needed⁵. Defects in the ion-transporting proteins in cells of the thick ascending limb lead to Bartter's syndrome⁶.

The energy for ion transport in cells of the thick ascending limb comes from the breakdown of ATP molecules — the cellular energy store — by the ubiquitous Na^+, K^+ -ATPase. This protein sits in the cell membrane on the external surface of the tubule (that is, on the 'basolateral' surface of the cell; Fig. 1). It uses energy from ATP to transport Na^+ ions out of the cell into the external tissue fluid, and K^+ ions into the cell from the tissue fluid. This reduces the intracellular Na^+ concentration and raises the intracellular K^+ concentration relative to that in the surrounding fluids.

The resulting concentration gradient of

Na^+ ions provides the driving force to import Na^+ , K^+ and two Cl^- ions from the fluid in the thick ascending limb, on an ion 'co-transporter' called NKCC2. In this way the intracellular Cl^- and K^+ levels are raised above electrochemical equilibrium and can leave the cell by diffusion through ion channels. K^+ ions recycle across both cell surfaces through channels, such as ROMK channels when recycling back into the thick ascending limb. Cl^- ions leave across the basolateral membrane through ClC-Kb channels. The result is the net transport of NaCl out of the thick ascending limb and into the tissue fluid for reabsorption. Mutations in NKCC2, ROMK or ClC-Kb can cause Bartter's syndrome⁶.

Surprisingly, the ion-transport processes of cells in the ear's stria vascularis⁷ are similar to this (Fig. 2). Here, intracellular ion concentrations comparable to those of the thick ascending limb are achieved by the concerted action of the Na^+, K^+ -ATPase and a different form of the co-transporter, NKCC1. Potassium ions then leave the cells through KCNQ1/KCNE1 channels, generating a high concentration of K^+ ions in one of the fluid-filled chambers of the ear (the endolymph). The high K^+ concentration aids in the mechanism whereby sound waves are converted to a voltage signal suitable for relay to the brain. Mutations in NKCC1, KCNE1 and KCNQ1 are all associated with deafness (see ref. 4 for references).

In both cases, Cl^- channels are vital: the accumulation of Cl^- ions within the cell cannot continue indefinitely because it would oppose further ion uptake, and all ion transport would cease. The channels that allow the efflux of Cl^- ions are members of the ClC-K family, which are expressed exclusively in the kidneys and ears. In kidneys, ClC-Ka is expressed in the thin ascending limb and ClC-Kb in the adjacent thick ascending limb. In the ear, both channel types are expressed in the stria vascularis, as Estévez *et al.* show⁴.

As mentioned above, mutations of ClC-Kb result in Bartter's syndrome. Defects in ClC-Ka in humans have not yet been reported, although disruption of the mouse counterpart causes renal insufficiency (progressive loss of kidney function) and diabetes insipidus (an inability to concentrate urine)⁸. But the absence of either ClC-Ka or ClC-Kb would not be expected to cause deafness, as the other channel would presumably compensate. Indeed, the ClC-Kb mutations that cause Bartter's syndrome lead only to kidney malfunction, not to deafness.

Presumably, however, the barttin protein³ has a pivotal role in both ears and kidneys, as it is mutated in people with type IV Bartter's syndrome. But what does it do? Estévez *et al.*⁴ aimed to find out. They showed that barttin is expressed together with ClC-K channels in certain mouse tissues, including

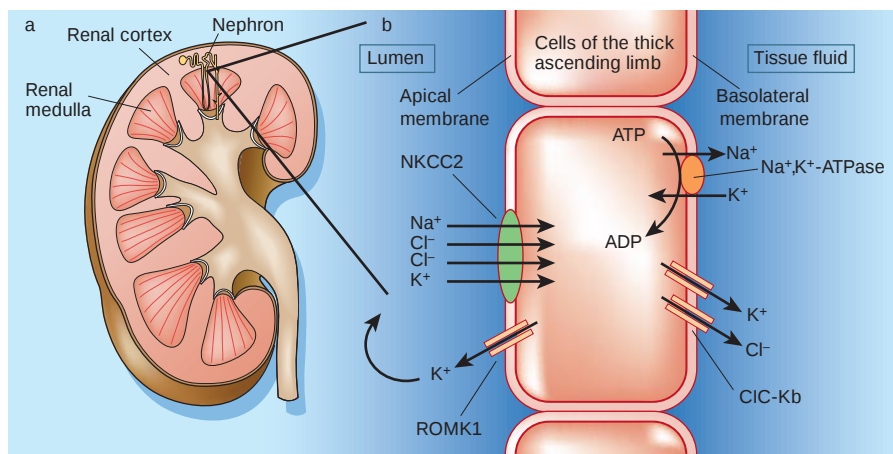


Figure 1 Ion transport in the kidney. **a**, The kidney. The renal cortex and medulla are made up of some half a million nephrons — the hollow tubules through which blood fluid flows to be filtered. One part of the nephron is the thick ascending limb. **b**, As fluid flows through the thick ascending limb, ions are extracted and returned to the body through the cells that form the tubule walls. Using energy supplied by hydrolysis of ATP, the Na^+, K^+ -ATPase (right-hand side) imports K^+ ions from the tissue fluid and exports Na^+ ions. The resulting concentration gradients allow another ion transporter, NKCC2 (left), to import Na^+ , Cl^- and K^+ ions from the fluid in the tubule. K^+ and Cl^- ions may then leave the cell by diffusing through relevant channels (including ROMK1 and ClC-Kb channels). K^+ ions merely recycle across the apical and basolateral membranes. But the net result is that Na^+ and Cl^- ions are concentrated in the tissue fluid and returned to the blood.

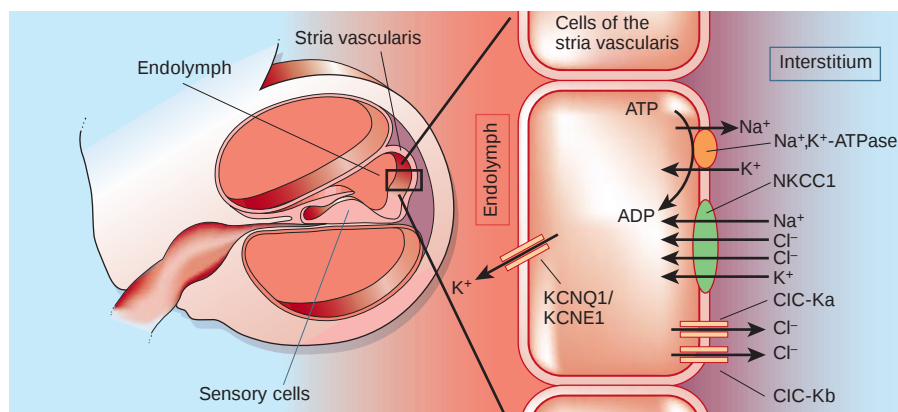


Figure 2 Ion transport in the inner ear. The balance of activities of various ion channels and ion transporters (including Na^+ , K^+ -ATPase, NKCC1, KCNQ1, KCNE1 and the CIC-K channels) ensures that the concentration of K^+ ions is high inside the endolymph of the ear. This aids in the mechanism by which sensory cells convert sound waves to voltage signals for transmission to the brain.

the stria vascularis, thin and thick ascending limbs, and other kidney cells that are known to express Cl^- channels. Moreover, barttin is needed for these channels to function: injecting RNA that encodes CIC-Ka or CIC-Kb into frog eggs did not result in measurable Cl^- currents, but also injecting RNA that encodes barttin resulted in Cl^- currents characteristic of those in ear or kidney cells. In addition, Estévez *et al.* introduced into barttin specific mutations that are seen in patients with type IV Bartter's syndrome. The mutations reduced the ability of the protein to support the expression of CIC-K channels. The authors' proposal that barttin is an accessory subunit (a β -subunit) of CIC-K channels seems the most reasonable explanation of these findings.

Barttin is the first known β -subunit of a Cl^- channel. But it may be that such subunits are the norm and that most ion channels

rely on them. Indeed, accessory subunits have been found to affect the expression and properties of several other channels, including various K^+ , Na^+ and Ca^{2+} channels⁹. The addition of β -subunits to the ion-channel armoury adds yet another layer of flexibility and diversity to ion-channel function. ■

Malcolm Hunter is at the School of Biomedical Sciences, Worsley Medical and Dental Building, University of Leeds, Leeds LS2 9NQ, UK.
e-mail: m.hunter@leeds.ac.uk

1. Bartter, F. C., Pronove, P., Gill, J. R. Jr & MacCardle, R. C. *J. Am. Soc. Nephrol.* **9**, 516–528 (1998).
2. Landau, D., Shalev, H., Ohaly, M. & Carmi, R. *Am. J. Med. Genet.* **59**, 454–459 (1995).
3. Birkenhager, R. *et al. Nature Genet.* **29**, 310–314 (2001).
4. Estévez, R. *et al. Nature* **414**, 558–561 (2001).
5. Greger, R. *Physiol. Rev.* **65**, 760–793 (1985).
6. Scheinman, S. J., Guay-Woodford, L. M., Thakker, R. V. & Warnock, D. G. *New Engl. J. Med.* **340**, 1177–1187 (1999).
7. Wangemann, P. H. & Schacht, J. in *The Cochlea* (eds Dallos, P., Popper, A. N. & Fry, R. R.) 130–185 (Springer, New York, 1996).
8. Matsumura, Y. *et al. Nature Genet.* **21**, 95–98 (1999).
9. Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, MA, 2001).

Materials science

A broader view of membranes

Roderic Lakes

Membranes that get fatter when they are stretched are considered counterintuitive, but may be more common than we think. They might even turn up in human tissue.

Stretch a rubber band and it becomes thinner; squeeze a rubber eraser and it becomes fatter. Most common, and not-so-common, materials share this property, even those that are so stiff you cannot see the deformation. The opposite effect — becoming fatter when stretched — is a property of some specially designed materials, notably foams with unusual structures, as well as certain single crystals. Writing in *Physical Review Letters*, Bowick and co-workers¹ show that a broad class of membrane structures also becomes fatter when

stretched. These structures include certain types of naturally occurring membranes, so if we look closely we may find that some biological cells deform in unexpected ways.

Elastic materials deform easily when under strain but return to their original dimensions when the force is removed — they resist changes to both shape and volume. Rubbery materials, which easily change shape but not volume, become notably thinner in cross-section when stretched. This is described by Poisson's ratio, the ratio of transverse contraction to longitudinal

extension during stretching. For rubber, Poisson's ratio is close to the theoretical upper limit of 0.5; for most other common materials it is between 0.25 and 0.35. Because all these materials become thinner when stretched, Poisson's ratio is always positive. The reason for the thinning is that interatomic bonds tend to align when deformed.

A negative Poisson's ratio, which requires a transverse expansion (thickening) on stretching, is considered counterintuitive; indeed such materials were once thought not to exist or even to be impossible. But materials with negative Poisson's ratio do occur², and have been called anti-rubber, auxetic or dilational. For example, two-dimensional honeycomb structures have been developed with 'inverted' cells^{3,4}, which unfold when stretched (Fig. 1, overleaf). (Regular honeycomb lattices, like most materials, have a positive Poisson's ratio.) Similarly, some foam materials with a three-dimensional microstructure of 'inwardly bulging' cells⁵ also get fatter when stretched. These foams and honeycombs have such unusual elastic behaviour because of non-uniform unfolding or deformation of the microstructure⁶. Such materials are usually tougher and more resilient than most conventional materials, and so would make good knee pads, seat cushions, biomaterials and air filters that are easily tuned or cleaned.

Bowick and co-workers¹ have shown that structures known as 'fixed connectivity' membranes have a negative Poisson's ratio. Such membranes are assumed to have fixed connections: their bonds do not break. They occur in some biological^{7,8} and synthetic⁹ systems, and are a two-dimensional counterpart to one-dimensional molecular chains of polymers¹⁰. Like polymers they have a fractal-like structure, but unlike polymers, which are always crumpled, polymerized membranes can undergo a phase transformation from a crumpled phase at high temperatures to a 'flat' but locally rough phase at low temperatures.

The negative Poisson's ratio in fixed-connectivity membranes¹¹ is due to changes in entropy⁸ (a measure of thermal disorder) that occur with deformation. Stretching the membrane tends to flatten out undulations arising from thermal fluctuations (Fig. 2, overleaf), producing expansion in two directions. This deformation, as with the honeycomb and foam, is non-uniform. But the honeycomb expands in the same plane as that in which its microstructure unfolds. By contrast, the thermally induced undulations of the polymerized membrane are perpendicular to the surface of the membrane. Some cubic single crystals¹² can also expand in one direction and constrict in another, although a polycrystalline version of the cubic material has 'normal' elasticity. In polycrystalline materials the elastic properties of the individual crystals are averaged,

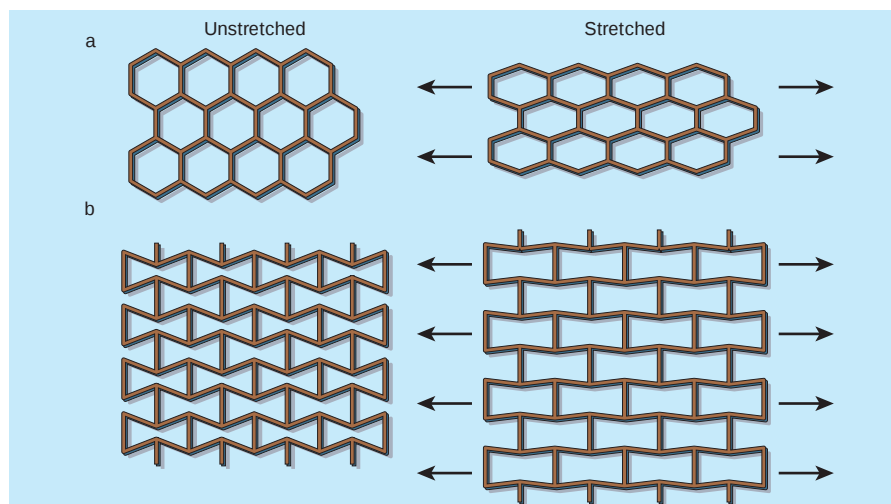


Figure 1 **Positive and negative Poisson's ratios.** Stretching these two-dimensional hexagonal structures horizontally reveals the physical origin of Poisson's ratio. **a.** The cells of regular honeycomb or hexagonal crystals elongate and narrow when stretched, causing lateral contraction and so a positive Poisson's ratio. **b.** In artificial honeycomb with inverted cells, the structural elements unfold, causing lateral expansion and a negative Poisson's ratio.

so the overall Poisson's ratio is almost always positive.

The elastic behaviour of the membranes studied by Bowick *et al.*¹ is said to be 'universal' because the authors require only a sparse set of assumptions to predict the Poisson's ratio. They start with a simple network of nodes, resembling a fishing net with fixed connections, which they model using a Monte Carlo simulation. Bowick *et al.* show that a negative Poisson's ratio is a universal property of such systems, whether the membrane is dominated by rigid bonds that resist bending or by 'self-avoiding' interatomic forces that prevent portions of the structure overlapping.

This unusual form of elasticity may also arise in biological processes because the membranes considered by Bowick *et al.* are similar to the protein skeletons⁸ of biological

membranes. The overall elasticity of a cell membrane results from both the protein skeleton and the high lipid content, but the relative contributions are not yet known. Even so, Bowick and colleagues' results are provocative. If our usual expectations about how things deform do not apply to biological membranes then we may need to reconsider the influence of membrane mechanics⁷ on the shape of cells, the formation of vesicles, and the deformation of cells during life processes. For example, red blood cells are routinely deformed when they pass through fine blood capillaries. As they deform, the membrane skeleton can unfold, which might help to transport large molecules or expose reactive chemical groups. Similar concepts may be used in the design of industrial membranes for responsive filtration or catalysis. Materials with specific Poisson's ratios are always useful — cork, for example, has a ratio of almost zero, making it ideal for sealing wine bottles. For membranes with a negative Poisson's ratio, the applications are sure to keep expanding. ■

Roderic Lakes is in the Department of Engineering Physics, Engineering Mechanics Program at the University of Wisconsin-Madison, 1500 Engineering Drive, Madison, Wisconsin 53706-1687, USA. e-mail: lakes@engr.wisc.edu

1. Bowick, M., Cacciuto, A., Thorleifsson, G. & Traverset, A. *Phys. Rev. Lett.* **87**, 148103 (2001).
2. <http://silver.neep.wisc.edu/~lakes/Poisson.html>
3. Kolpakov, A. G. *Prikl. Mat. Mekh.* **59**, 969–977 (1985).
4. Gibson, L. J., Ashby, M. F., Schajer, G. S. & Robertson, C. I. *Proc. R. Soc. Lond. A* **382**, 25–42 (1982).
5. Lakes, R. S. *Science* **235**, 1038–1040 (1987).
6. Lakes, R. S. *J. Mat. Sci.* **26**, 2287–2292 (1991).
7. Lipowsky, R. *Nature* **349**, 475–481 (1991).
8. Schmidt, C. F. *et al. Science* **259**, 952–955 (1993).
9. Wen, X. *et al. Nature* **355**, 426–428 (1992).
10. Kantor, Y., Kardar, M. & Nelson, D. R. *Phys. Rev. Lett.* **57**, 791–794 (1986).
11. Aronovitz, J. A. & Lubensky, T. C. *Phys. Rev. Lett.* **60**, 2634–2637 (1988).
12. Baughman, R. H., Shacklette, J. M., Zakhidov, A. A. & Stafstrom, S. *Nature* **392**, 362–365 (1998).

Daedalus

The art of copying

The libraries of the ancient world survived only by continual copying, and the slow crumbling of paper threatens ours as well. So Daedalus wants to be able to copy a closed book at speed. Ideally, the librarian should be able simply to load a block of books onto the copier, which would then transfer their contents to compact disk.

A paper page is about 0.1 mm thick, a typical infrared wavelength. It is composed of fibres of much finer diameter, so it should be pretty transparent to light with a wavelength of 0.1 mm or less. At such a wavelength, each page should be optically distinguishable from the pages on either side. A pulsed infrared laser beam launched perpendicularly into the side, top or bottom of the book, or into all three simultaneously, should illuminate one page only, and would follow it even if the page was a little bent.

A sufficiently narrow laser beam can trace a single line of type, or pick out single letters from successive lines. Each letter is an opaque indentation of the paper; it will absorb the light pulse a little, and this sharp attenuation will return a reflection, or 'echo', characteristic of the shape of the letter. A beam-splitter could then separate the stream of returning echoes into a photocell for computer analysis.

Unfortunately, a page is printed on both sides. At any printed site, the paper may be indented by two opaque letters, both of which will contribute to the echo. But this is where beams from all three sides come in. Each could be swung through an angle to illuminate any of several files of letters. Cunning software with linguistic knowledge will be needed to resolve the multiple ambiguities. But a closed page lit by pulsed lasers should be readable.

DREADCO physicists and librarians are now at work. Their prototype steps through a block of books at high speed, a page at a time. The final device will probably send its output to a compact disk, archival microfilm or miniprint on alkaline paper. Photographs and diagrams are a problem, but they could be identified by the software, and the book opened and photographed the hard way.

The new DREADCO book-copier will transform the task faced by the world's librarians. Instead of having to open every book at every page, they will merely have to load books in bulk into the copier. Its output will be far smaller, neater and more easily stored. The mighty achievements of the nineteenth and twentieth centuries will be faithfully recorded for future historians.

David Jones

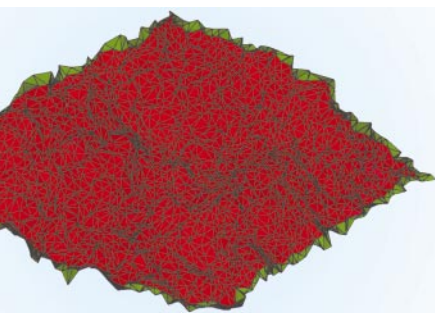


Figure 2 **An artificial membrane with a negative Poisson's ratio.** The local roughness of a 'flat' fixed-connectivity membrane is due to thermal fluctuations. Stretching the membrane flattens the roughness, causing lateral expansion in the plane of the membrane. In their simulation, Bowick *et al.*¹ find that all fixed-connectivity membranes, possibly including some biological ones, share this property. (After M. Bowick, with permission.)

Asymmetric Broca's area in great apes

A region of the ape brain is uncannily similar to one linked with speech in humans.

Brodman's area 44 delineates part of Broca's area within the inferior frontal gyrus of the human brain and is a critical region for speech production^{1,2}, being larger in the left hemisphere than in the right¹⁻⁴ — an asymmetry that has been correlated with language dominance^{2,3}. Here we show that there is a similar asymmetry in this area, also with left-hemisphere dominance, in three great ape species (*Pan troglodytes*, *Pan paniscus* and *Gorilla gorilla*). Our findings suggest that the neuroanatomical substrates for left-hemisphere dominance in speech production were evident at least five million years ago and are not unique to hominid evolution.

To our knowledge, no one has assessed whether there is any consistent left-right anatomical asymmetry in the inferior frontal gyrus (IFG) of any non-human primate. This is surprising because it is known from cytoarchitectonic and electrical stimulation studies that many non-human primates, including the great apes^{5,6}, possess a homologue of area 44. We have therefore investigated whether homologous neuro-anatomical asymmetries are present in area 44 in great apes.

From magnetic resonance images (MRI) obtained from 20 chimpanzees (*P. troglodytes*), 5 bonobos (*P. paniscus*) and 2 gorillas (*G. gorilla*; Fig. 1a), we found that area 44 in these species shows a pattern of morphological asymmetry that has left-hemisphere surface-area predominance similar to the homologous cortical area of humans (mean left, $127.7 \text{ mm}^2 \pm 8.1 \text{ s.e.}$; mean right, $104.2 \text{ mm}^2 \pm 6.1 \text{ s.e.}$; $F(1,25) = 7.45$, $P = 0.011$; see supplementary information for details of the statistical analysis, and see Fig. 1b for individual asymmetry measures). In humans, this region is part of Broca's area, a key anatomical substrate for speech functions, particularly in motor aspects of speech such as articulation and fluency^{2,7}.

The part possession by great apes of a homologue of Broca's area is puzzling, particularly considering the discrepancy between sophisticated human speech and the primitive vocalizations of great apes. This may be explained by the contribution that gestures have made to the evolution of human language and speech⁸. In monkeys, the so-called 'mirror neurons' in area 44 seem to subserve the imitation of hand

grasping and manipulation, and this neural system may have been specialized initially for gestural and later for vocal communication⁹. In captive great apes, manual gestures are both referential and intentional^{10,11}, and are preferentially produced by the right hand (which is controlled by the left hemisphere). This right-hand bias is consistently greater when gesturing is accompanied by vocalization¹⁰.

From an evolutionary standpoint, therefore, asymmetry in area 44 may be associated with the production of gestures accompanied by vocalizations in great apes, an ability that eventually selected for the development of speech systems in modern humans and perhaps generated more cortical folding in the IFG, leading to expansion of Brodmann's area 45 in the human brain.

Whatever the function of area 44 in great apes, our finding that these species show a human-like asymmetry not only in posterior (such as the planum temporale^{12,13}) but also in frontal regions, indicates that the origin of asymmetry in language-related areas of the human brain should be interpreted in evolutionary terms rather than being confined to the human species.

Claudio Cantalupo*†, William D. Hopkins*‡†

*Division of Psychobiology and Living Links Center, Yerkes Regional Primate Research Center, Emory University, Atlanta, Georgia 30322, USA
e-mail: lrcbh@rmy.emory.edu

†Language Research Center, Georgia State University, Atlanta, Georgia 30303, USA

‡Department of Psychology, Berry College, Mount Berry, Georgia 30149, USA

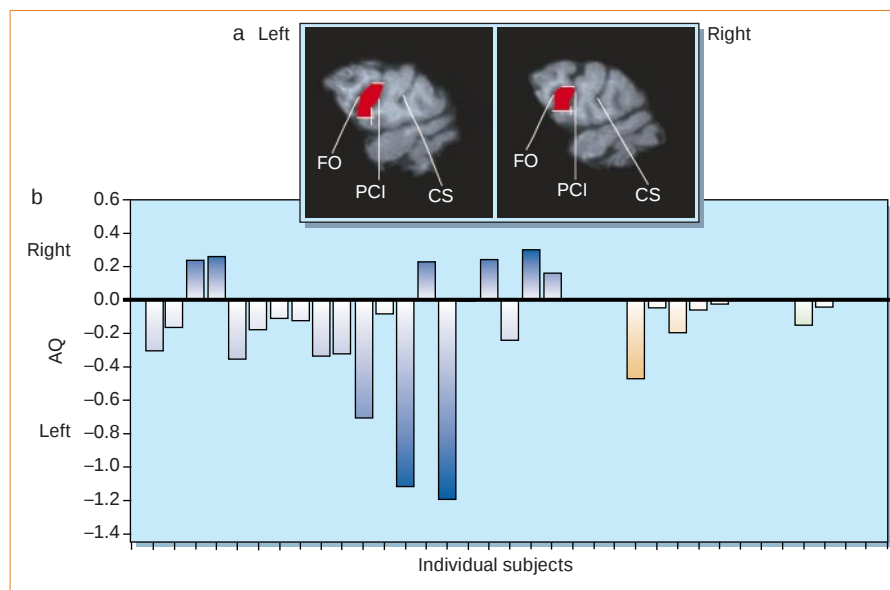


Figure 1 Asymmetry of Brodmann's area 44 in great apes. **a**, Area 44 (shown in red in the two hemispheres of a chimpanzee brain), which is defined as the portion of cortical surface of the inferior frontal gyrus (IFG) bounded by the fronto-orbital (FO) and precentral-inferior (PCI) sulci (CS, central sulcus). Both the FO and PCI sulci can be seen in parasagittal (1 mm thick) slices obtained using magnetic resonance imaging. The dorsal and ventral borders of area 44 are marked in white. We base the definition of the ventral border on cytoarchitectonic studies that show area 44 expanding over the ventral extent of the IFG, even if the PCI sulcus usually terminates at a more dorsal level^{5,6}. **b**, Individual asymmetry quotients (AQ; calculated as $(R - L) / [(R + L) \times 0.5]$, where R and L represent areas in the right and left regions, respectively; the sign of the quotient indicates the direction of asymmetry: positive, right-hemisphere asymmetry; negative, left-hemisphere asymmetry) in individual chimpanzees (blue bars), bonobos (orange) and gorillas (green). On the basis of criteria used for humans³, 20 apes showed left-hemisphere asymmetry, six showed right-hemisphere asymmetry and one had no asymmetry ($\chi^2(2) = 21.56$, $P < 0.001$). The number of subjects with a left-hemisphere asymmetry was greater than the number of subjects with either right-hemisphere asymmetry ($z = 2.55$, $P = 0.011$) or no asymmetry ($z = 4.15$, $P < 0.001$).

- Galaburda, A. M. in *Cerebral Dominance: The Biological Foundations* (eds Geschwind, N. & Galaburda, A. M.) 11–25 (Harvard Univ. Press, Cambridge, Massachusetts, 1984).
- Foundas, A. L., Eure, K. F., Luevano, L. F. & Weinberger, D. R. *Brain Lang.* **64**, 282–296 (1998).
- Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B. & Zilles, K. *J. Comp. Neurol.* **412**, 319–341 (1999).
- Broca, P. *Bull. Soc. Anat. Paris* **2**, 330–357 (1861).
- von Bonin, G. in *The Precentral Motor Cortex* (ed. Bucy, P. C.) 7–82 (Univ. Illinois Press, Urbana-Champaign, 1949).
- Bailey, P., von Bonin, G. & McCulloch, W. S. *The Isocortex of the Chimpanzee* (Univ. Illinois Press, Urbana-Champaign, 1950).
- Luria, A. R. *Traumatic Aphasia* (Mouton, The Hague, 1970).
- Corballis, M. C. *Cognition* **44**, 197–226 (1992).
- Rizzolatti, G. & Arbib, M. A. *Trends Neurosci.* **21**, 188–194 (1998).
- Hopkins, W. D. & Leavens, D. A. *J. Comp. Psychol.* **112**, 95–99 (1998).
- Hostetter, A. B., Cantero, M. & Hopkins, W. D. *J. Comp. Psychol.* (in the press).
- Gannon, P. J., Holloway, R. L., Broadfield, D. C. & Braun, A. R. *Science* **279**, 220–222 (1998).
- Hopkins, W. D., Marino, L., Rilling, J. K. & MacGregor, L. A. *NeuroReport* **9**, 2913–2918 (1998).

Supplementary information accompanies this communication on Nature's website (www.nature.com).

Entomology

Immune defence in bumble-bee offspring

Immune-challenged vertebrate females transfer specific antibodies to their offspring^{1–3}, but this gratuitous immunity cannot operate in invertebrates⁴. Here we show that constitutive immune defence is enhanced in sexual offspring of the bumble-bee *Bombus terrestris* L. when the parental colony is immune-challenged. Our findings indicate that invertebrates may use a different component of the immune system to generate a facultative trans-generational increase in the immune response.

Insect immunity is characterized by the inducible expression of a large array of antimicrobial peptides and by the constitutive melanization–encapsulation response, which is based on a cascade involving an inactive precursor of the enzyme phenol oxidase^{4,5}. Antibacterial activity can be induced, for example, by lipopolysaccharide (LPS) extracted from bacterial surfaces. The operation of the cascade is indicated by the phenol oxidase activity in the insect haemolymph^{6,7} and can be monitored by measuring the rate of conversion of a phenol substrate into quinone, which then polymerizes to form melanin. Because both quinone and melanin are toxic to microorganisms⁵, hosts with high phenol oxidase activity are less susceptible to microbial infection⁸.

Social insects cooperate in brood care and make a considerable investment in their offspring. In annual species such as bumble-bees, reproduction occurs at the end of the colony cycle, when sexuals (daughter queens and males) emerge — here the term ‘trans-generational’ distinguishes the queen and workers from sexual offspring. Unlike daughter queens, males do not hibernate, so their reproductive success depends on post-emergence survival after they leave the parental colony and are exposed to parasites in the same habitat⁹. Assuming facultative adjustment of offspring immunity, we investigated whether parasite-challenged parental colonies could enhance their males’ immunocompetence.

We used a split-colony design¹⁰ with 11 colonies, each equally split into treatment and control groups. In the immune-challenged group, 70–80% of workers were injected weekly with LPS (Sigma L-2755, 0.5 mg ml^{−1} in Ringer’s solution (5 µl)), which activates the immune system for long periods¹¹. Control workers were treated in the same way, but with the omission of LPS. Colonies completed their life cycle in the laboratory under standard conditions (24 °C, 60% relative humidity). We counted the number of sexuals and haemocytes (Neubauer haemocytometer, 1/6 dilution) and used standard protocols to measure

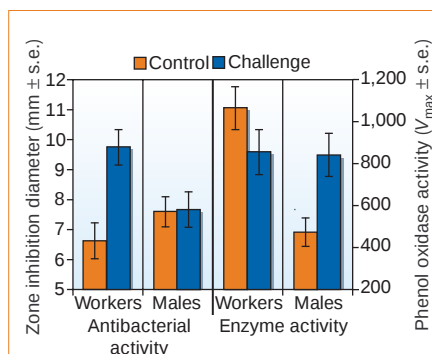


Figure 1 Antibacterial and phenol oxidase activities in the haemolymph of workers and males from control (orange bars) and challenged (blue bars) groups from 11 bumble-bee colonies. In workers, antibacterial activity was higher in challenged groups than in controls (Wilcoxon’s paired signed-rank test: $T_+ = 6$, $n = 11$, $P < 0.02$), but phenol oxidase activity was lower ($T_- = 3$, $n = 11$, $P = 0.005$). In males, antibacterial activity was the same ($T_+ = 31$, $n = 11$, NS) but phenol oxidase activity was higher ($T_+ = 6$, $n = 11$, $P < 0.02$) in the challenged side than in controls. Further experiments showed that increased activity in males correlates with an increased encapsulation response against an invader ($F_{1,20} = 5.25$, $P = 0.031$). V_{max} is measured as the maximum change in optical density $\times 10^{-3}$ per minute.

antibacterial¹² and phenol oxidase¹³ activity (1/20 dilution).

As expected, workers in the challenged groups showed more antibacterial activity than controls (Fig. 1). Their phenol oxidase activity, however, was lower (Fig. 1), indicating that there could be a possible trade-off between these two immune responses in challenged workers. Haemocyte counts were similar between the two groups (Wilcoxon’s paired signed-rank test, $T_- = 29$, $n = 11$, NS). Immune-challenged groups had lower reproductive output (repeated measures-MANOVA for log-transformed number of males and queens: Hotelling’s $T = 1.297$, $F_{2,9} = 5.839$, $P = 0.024$), notably producing fewer queens ($F_{1,10} = 12.082$, $P = 0.006$), indicating a possible trade-off between reproductive output and immune response. Male offspring from challenged groups showed higher phenol oxidase activity than controls, but antibacterial activity (Fig. 1) and haemocyte counts were comparable between the two groups ($T_- = 29$, $n = 11$, NS).

As insects do not produce antibodies, they cannot transfer specific immunity as mammals do¹. Male bumble-bees from immune-challenged groups have increased constitutive immunity relative to controls, which both enhances encapsulation (Fig. 1) and protects against microorganisms^{5,8}. As the phenol oxidase enzyme cascade provides a broader immunity than the costly antibacterial immune response¹², males may benefit by enhancing their most general means of prophylaxis. Although the physiological mechanism by which this trans-generational transfer is achieved is unknown, the enhanced immunity could be

the result of monitoring cues from worker bees, as in the density-dependent prophylaxis observed in other insects¹³.

Yannick Moret, Paul Schmid-Hempel

Eidgenössische Technische Hochschule Zürich, Ecology and Evolution, ETH-Zentrum NW, 8092 Zürich, Switzerland

e-mail: moret@eco.unm.w.ETH.ZH.CH

- Hanson, L. A. *Ann. Allergy Asthma Immunol.* **81**, 523–537 (1998).
- Coste, A., Sirard, J. C., Johansen, K., Cohen, J. & Kraehenbuhl, J. P. *J. Virol.* **74**, 8966–8971 (2000).
- Rollier, C., Charolles, C., Jamard, C., Trepo, C. & Cova, L. *Vaccine* **18**, 3091–3096 (2000).
- Hoffmann, J. A., Reichhart, J. M. & Hetru, C. *Curr. Opin. Immunol.* **8**, 8–13 (1996).
- Söderhäll, K. & Cerenius, L. *Curr. Opin. Immunol.* **10**, 23–28 (1998).
- Nappi, A. J. & Vass, E. *Pigment Cell Res.* **6**, 117–126 (1993).
- Thörnqvist, P.-O. & Söderhäll, K. in *Diseases in Asian Aquaculture* vol. 3 (eds Flegel, T. W. & MacRae, I. H.) 203–218 (Fish Health Section, Asian Fisheries Soc., Manila, Philippines, 1996).
- Nigam, Y., Maudlin, I., Welburn, S. & Ratcliffe, N. A. *J. Invertebr. Pathol.* **69**, 279–281 (1997).
- Schmid-Hempel, P. *Parasites in Social Insects* (Princeton Univ. Press, Princeton, New Jersey, 1998).
- Schmid-Hempel, R. & Schmid-Hempel, P. *Funct. Ecol.* **12**, 22–30 (1998).
- Kato, Y. *et al.* *Insect Biochem. Mol. Biol.* **24**, 539–545 (1994).
- Moret, Y. & Schmid-Hempel, P. *Science* **290**, 1166–1168 (2000).
- Barnes, A. I. & Siva-Jothy, M. T. *Proc. R. Soc. Lond. B* **267**, 177–182 (2000).

Nanotechnology

Synthesis of carbon ‘onions’ in water

The fabrication of carbon nanomaterials usually calls for expensive vacuum systems to generate plasmas^{1,2} and yields are disappointingly low. Here we describe a simple method for producing high-quality spherical carbon nano-‘onions’ in large quantities without the use of vacuum equipment. The nanoparticles, which have C₆₀ cores surrounded by onion-like nested particles, are generated by an arc discharge between two graphite electrodes submerged in water. This technique is economical and environmentally benign, and produces uncontaminated nanoparticles which may be useful in many applications.

Nanoparticles have previously been prepared using a range of vacuum and non-vacuum methods^{3–5}. Ours is a non-vacuum method in which the carbon arc is sustained in deionized water. The apparatus consists of two submerged graphite electrodes, and the arc discharge is initiated by contacting a pure grounded graphite anode (tip diameter, 5 mm) with the carbon cathode (tip diameter, 12 mm) of similar purity; the discharge voltage and current were 16–17 V and 30 A, respectively. The nano-onions are mostly found floating on the water surface, with the rest falling to the bottom of the beaker through natural segregation, giving material of high purity.

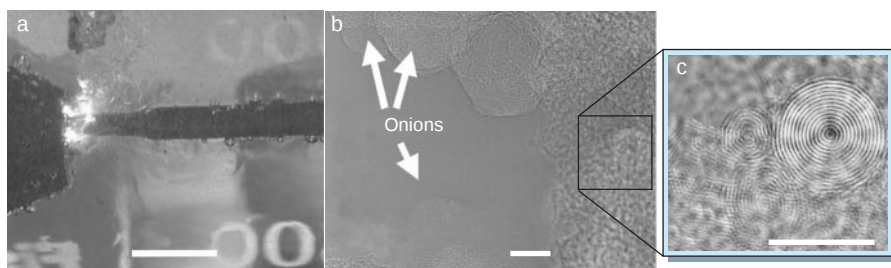


Figure 1 Carbon nano-'onions' created by arc discharge in water. **a**, Image of a carbon arc discharge in water. Scale bar, 12 mm. **b, c**, Low- and high-magnification electron micrographs of carbon nano-onions floating on the water surface after their production. Scale bars, 10 nm.

A digital image of the arc discharge in water is shown in Fig. 1a.

We investigated the unpurified material from the water surface using a JEOL 4000EX transmission electron microscope. A typical high-resolution micrograph of the material is shown in Fig. 1b — several spherical carbon nano-onions are evident, as well as polyhedral, nested onion-like particles. At higher magnification (Fig. 1c), spherical nano-onions with 7, 10 or 15 walls are seen. The core of the larger 'onion' in Fig. 1c has a diameter of 7–8 Å, which is consistent with that of the C_{60} molecule. A series of images obtained at intervals through the features verified that they were spherical nanoparticles and not nanotubes⁶.

The average diameter of the nano-onions is 25–30 nm (range 5–40 nm), a useful size range for many lubrication applications. Nanoparticles composed of halogens, tungsten and sulphur (IF- WS_2), which are similar to the carbon onions reported here, are more effective as solid-state lubricants when dispersed in oil than are the conventionally used 2H-MoS₂ crystals⁷.

Although the production rate (3 mg min⁻¹) of these carbon onions is faster than in conventional processes, it is not yet adequate for industrial application. The apparatus used in the initial experiments was unsophisticated, however, and minor adjustments (such as increasing the arc current to 50 A) could improve yields significantly. The loss of water through evaporation is slight, but chilling and circulating it may enhance production; auto-feeding the carbon anode would allow the process to run continuously for several hours. With these and other minor modifications, our process should be adaptable for mass production of nanoparticles.

N. Sano*†, H. Wang*, M. Chhowalla*, I. Alexandrou*, G. A. J. Amaratunga*

*Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK

e-mail: mc209@eng.cam.ac.uk

†Present address: Department of Chemical Engineering, Himeji Institute of Technology, Himeji 671-2201, Japan

3. Cabiocch, T., Riviere, J. P. & Delafond, J. J. *Mater. Res.* **30**, 4787–4792 (1995).
4. Ishigami, M., Cumings, J., Zettl, A. & Chen, S. *Chem. Phys. Lett.* **319**, 457–459 (2000).
5. Hsu, W. K. et al. *Nature* **377**, 687 (1995).
6. Hsin, Y. L. et al. *Adv. Mater.* **13**, 830–833 (2001).
7. Tenne, R., Homyonfer, M. & Feldman, Y. *Chem. Mater.* **10**, 3225–3238 (1998).

COMMUNICATIONS ARISING

Palaeoecology

Fossils and avian evolution

Discoveries of archaic Mesozoic fossil birds ('opposite' birds, or enantiornithines) during the past decade have revolutionized our understanding of early avian evolution, but the rarity of Early Cretaceous ornithurines — birds that are closely related to the modern avian radiation — has meant that information about these species has lagged behind.

Norell and Clarke¹ describe the newly discovered and well preserved Late Cretaceous Mongolian *Apsaravis* as an ornithurine, cladistically slotting it between the well known and abundant marine hesperornithiforms and ichthyornithiforms. They claim that this specimen provides evidence against a so-called near-shore or marine 'ecological bottleneck' of ornithurines at the Cretaceous/Tertiary (K/T) boundary², as well as insight into the assembly of the modern flight apparatus. However, I question their conclusions on the grounds of their inconsistency with literature on earlier ornithurine birds and of comparison of *Apsaravis* with other abundant ornithurines of that age.

Contrary to the claims of Norell and Clarke¹, morphologically well defined ornithurines are known from highly preserved fossil material from the Early to the latest Cretaceous period, including *Ambiortus* from the Early Cretaceous of Mongolia (which shares no particular adaptations for near-shore habitats and has a highly developed flight apparatus with a deeply keeled ornithurine sternum) and the closely allied *Otogornis* (Early Cretaceous of China). Other notable examples include *Chaoyangia*, a toothed ornithurine from the Early Cretaceous of continental China³, the Early

Cretaceous *Gansus* (possibly a shore-dwelling bird), from Gansu Province in the continental interior of China, and *Liaoningornis*, perhaps the oldest known ornithurine bird from the famous Early Cretaceous *Confuciusornis* locality².

As hesperornithiform birds (highly derived ornithurines) were already well adapted for foot-propelled diving by the Aptian epoch (the latter part of the Early Cretaceous), their divergence must have been from even earlier ornithurine birds. By the Late Cretaceous, numerous taxa of hesperornithiform birds, together with the volant ornithurines *Ichthyornis* and *Apatornis*, are known from many marine localities in both the Northern and Southern Hemispheres.

What correlation is there between the distributions of shore birds (waders; Charadriiformes) and the marine, shoreline habitats of the continental margins? These birds are quite cosmopolitan, occurring in shallow waters of continental margins and interiors, and in freshwater and marine habitats.

As *Apsaravis* is not related to any modern order of birds, it has no bearing on what types of ornithurine bird breached the K/T boundary. It has been noted previously^{2,4} that no modern avian order (except Charadriiformes) can be confidently dated to before the K/T boundary⁵. The enantiornithines were the predominant land birds of the Mesozoic era, identified from numerous localities from the Lower to the Upper Cretaceous, in contrast to the scant finds of ornithurines; no enantiornithine is known to have lived after the K/T boundary. Hesperornithiforms from the Lower to Upper Cretaceous and ichthyornithiforms from the Upper Cretaceous have been found from many marine localities, but not after the K/T boundary; early Palaeocene fossil birds are known predominantly from morphotypes representative of shore birds.

The K/T extinction event was therefore likely to have been devastating for birds. 'Shore birds' may represent one of the major bottlenecks of avian morphotypes, transcending the K/T boundary, and could represent the wellsprings of an explosive Tertiary radiation of the modern bird orders that closely paralleled a similar event in Tertiary mammal history (this explosive model is supported by 'gap' analyses^{6,7}).

Because shoreline habitats were as common in continental interiors as on continental margins during the Late Cretaceous², and shore birds are equally at home in freshwater and/or marine environments, *Apsaravis* has no bearing on the habitat of Late Cretaceous birds. If, for example, an individual of *Larus pipixcan*, the common gull of summer prairies of the American great plains, were to be fossilized, the logic of Norell and Clarke would imply that it was a member of a land-bird order.

1. Kratschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. *Nature* **347**, 354–357 (1990).

2. Thess, A. et al. *Science* **273**, 483–487 (1996).

However, Mongolia was characterized in the Late Cretaceous by extensive lakes, possibly with marine connections, and by the Campanian it would perhaps be most accurate to describe it as a desert, as in South Africa, with a swampy inland delta⁸. Small, possibly volant hesperornithiforms and *Presbyornis*, a widespread wader with webbed feet, have been found at nearby sites of about the same age⁹.

I consider *Apsaravis* to have little to contribute to our understanding of avian evolution, and its lack of a clear relationship with any kind of modern bird makes its significance ambiguous. If *Apsaravis* is not related to any modern ornithurine, how can it tell us anything important about the evolutionary questions raised by Norell and Clarke?

Alan Feduccia

Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599-3280, USA
e-mail: feduccia@bio.unc.edu

1. Norell, M. A. & Clarke, J. A. *Nature* **409**, 181–184 (2001).
2. Feduccia, A. *The Origin and Evolution of Birds* 2nd edn (Yale Univ. Press, New Haven, Connecticut, 1999).
3. Hou, L.-H., Martin, L. D., Zhou, Z. H. & Feduccia, A. *Science* **274**, 1164–1167 (1996).
4. Feduccia, A. *Science* **267**, 637–638 (1995).
5. Dyke, G. J. & Mayr, G. *Nature* **399**, 317–318 (1999).
6. Bleiweiss, R. *Geology* **26**, 323–326 (1998).
7. Marshall, C. R. *Geology* **27**, 95–96 (1999).
8. Shuvalov, V. F. in *The Cretaceous Stratigraphy and Palaeobiogeography of Mongolia* (eds Benton, M. J. et al.) 256–278 (Cambridge Univ. Press, Cambridge, 2000).
9. Kurochkin, E. N. in *The Cretaceous Stratigraphy and Paleobiology of Mongolia* (eds Benton, M. J. et al.) 533–559 (Cambridge Univ. Press, Cambridge, 2000).

Norell and Clarke reply — Given that Feduccia has explicitly stated that there is a near absence of ornithurine birds in Late Cretaceous continental deposits¹ and has speculated that ornithurines may have been more or less restricted to shoreline and marine deposits during this time¹, we do not believe that we misrepresented Feduccia's hypothesis. We reported the finding of an almost complete skeleton of an ornithurine from Late Cretaceous continental deposits, and do not see how this specimen could have no bearing on Feduccia's previous arguments.

Feduccia comments that other Mesozoic bird specimens are more, or just as, useful for tackling questions concerning the origin of extant bird lineages. Although all specimens contain some information, we disagree with Feduccia's current assertion that *Apsaravis* is simply one of a group of "abundant" ornithurine fossils. The specimens he mentions are either not ornithurine or are so poorly preserved that they have not shed much light on their own phylogenetic positions, let alone on broader patterns of avian evolution.

The two "ornithurine" birds *Liaoningornis*

and *Chaoyangia* fall outside Ornithurae in Feduccia's own work. The explicit cladistic definition of Ornithurae (most recent common ancestor of Hesperornithiformes plus Aves and all descendants²) is less inclusive than Feduccia's more subjective definition (taxa other than those that are not 'modern' enough to be ornithurine). Feduccia's 'Ornithurae' is predicated on the existence of a 'Sauriurae' or the paraphyletic group that contains these primitive taxa. The fact that *Apsaravis* and our analyses add to the mounting evidence^{3,4} against sauriurine monophyly has been overlooked in Feduccia's estimation of the importance of *Apsaravis*.

Other specimens that we did not consider are problematic and underscore the importance of well preserved and phylogenetically placed taxa such as *Apsaravis*. Feduccia did not include the fragmentary *Otogornis*, *Ambiortus* and *Gansus* in his analyses⁵. Although he claims that *Chaoyangia* possesses a "toothed skull", the holotype actually consists only of a torso and partial hindlimbs. The "toothed skull" belongs to a specimen once referred to as *Chaoyangia*⁵ but later identified as the holotype of *Songlingornis linghensis*⁶. This specimen cannot be referred to *Chaoyangia* (as indeed it has not been⁶) as no element known from the holotype is also represented in the referred specimen.

Although we did not comment on the implications of *Apsaravis* for the timing of the origin of Aves, Feduccia's conjecture that it cannot inform our understanding of this origin (because it is not part of an extant lineage) is incompatible with his own arguments. In recounting the origin of Aves, he invokes taxa such as ichthyornithiforms and hesperornithiforms, which are not parts of extant lineages. Furthermore, it has been argued⁷ that gap analyses may be consistent with Cretaceous or Tertiary diversification of avian lineages, depending on what model of diversification rate and recovery potential is considered realistic.

Reasoning derived from phylogenetic analysis is a powerful way to test hypotheses of relationships or the evolution of morphology (for example, enantiornithine monophyly and novelties in the flight apparatus). We used a phylogenetic test to assess the idea that transitional 'shore birds' gave rise to all extant birds through an ecological bottleneck¹.

If such a bottleneck occurred, then when ecology is bracketed phylogenetically for living birds, 'shore bird' morphology and ecology should be basal to the crown clade, as well as in its nearest sister taxa. However, virtually all molecular and morphological evidence places 'land birds' (tinamous, ratites, galliforms and anseriforms, for example) at the base of Aves^{8,9}. Charadriiformes, the extant lineage referred to as

shore birds, are placed as derived forms within Aves^{8,9}. Thus, if the ecologies that are basal to the crown clade are bracketed, no support is found for such a bottleneck.

Apsaravis, because of its phylogenetic placement, constrains the inference of the ecologies of the most recent common ancestor of the avian crown clade. We do not understand how an ornithurine with no 'shore bird' morphologies, from a dune-field¹⁰, can be interpreted as compatible with Feduccia's idea¹ of ecological restriction of these taxa to shorelines and marine environments. If *Apsaravis* can simply be assumed, without consideration of phylogenetic tests, to have flown from an unknown nearby lake, then we do not see how Feduccia's hypothesis is testable.

Julia A. Clarke*, Mark A. Norell†

*Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06520-8109, USA
e-mail: julia.clarke@yale.edu

†Division of Paleontology, American Museum of Natural History, 79th Street at Central Park West, New York, New York 10024-5192, USA

1. Feduccia, A. *The Origin and Evolution of Birds* 2nd edn (Yale Univ. Press, New Haven, Connecticut, 1999).
2. Chiappe, L. M. *Alcheringa* **15**, 333–338 (1991).
3. Chiappe, L. M. *Nature* **378**, 349–355 (1995).
4. Forster, C. A., Chiappe, L. M. & Sampson, S. D. *Science* **382**, 532–534 (1996).
5. Hou, L.-H., Martin, L. D., Zhou, Z.-H. & Feduccia, A. *Science* **274**, 1164–1167 (1996).
6. Hou, L.-H. *Mesozoic Birds of China* (Taiwan Provincial Feng Huang Ku Bird Park, Nan Tou, 1997).
7. Marshall, C. R. *Geology* **27**, 95–96 (1999).
8. Groth, J. G. & Barrowclough, G. F. *Mol. Phylog. Evol.* **12**, 115–123 (1999).
9. Cracraft, J. & Clarke, J. A. in *Perspectives on the Origin and Early Evolution of Birds: Proceedings of the International Symposium in Honor of John H. Ostrom* (Peabody Mus. Nat. Hist., New Haven, Connecticut, in the press).
10. Loope, D. B., Dingus, L., Swisher, C. C. & Minjin, C. *Geology* **26**, 27–30 (1998).

retraction

Furtive mating in female chimpanzees

Pascal Gagneux, David S. Woodruff & Christophe Boesch
Nature **387**, 358–359 (1997).

In this genetic analysis of a community of chimpanzees in the Tai forest, Côte d'Ivoire (carried out in 1994), we concluded that 7 out of 13 offspring were sired by males not found in the mother's social group. Now a study of paternity using quantified and automated methods shows that the incidence of extra-group paternity is much lower (1 out of 14 offspring; ref. 1). Direct comparison at the only satellite locus re-examined reveals that 10 out of 66 alleles (15%) and 9 out of 33 individuals (27%) were inaccurately genotyped. Possible sources of error in the first study include allelic dropout in the amplification of degraded DNA from field-collected samples of shed hair, inconsistent visual autoradiograph interpretation (stutter bands), contamination and sample mix-up. The new analysis confirms that extra-group paternity can occur in nature, but shows that the social community probably corresponds to the reproductive unit in chimpanzees.

1. Vigilant, L., Hofreiter, M., Siedel, H. & Boesch, C. *Proc. Natl Acad. Sci. USA* **98**, 12890–12895 (2001).

Attosecond metrology

M. Hentschel^{*†}, R. Kienberger^{*†}, Ch. Spielmann^{*}, G. A. Reider^{*}, N. Milosevic^{*}, T. Brabec^{*}, P. Corkum[‡], U. Heinzmann[§], M. Drescher[§] & F. Krausz^{*}

^{*} Institut für Photonik, Technische Universität Wien, Gusshausstr. 27, A-1040 Wien, Austria

[‡] Steacie Institute of Molecular Sciences, NRC Canada, Ottawa, Canada K1A 0R6

[§] Fakultät für Physik, Universität Bielefeld, D-33615 Bielefeld, Germany

[†] These authors contributed equally to this work

The generation of ultrashort pulses is a key to exploring the dynamic behaviour of matter on ever-shorter timescales. Recent developments have pushed the duration of laser pulses close to its natural limit—the wave cycle, which lasts somewhat longer than one femtosecond (1 fs = 10⁻¹⁵ s) in the visible spectral range. Time-resolved measurements with these pulses are able to trace dynamics of molecular structure, but fail to capture electronic processes occurring on an attosecond (1 as = 10⁻¹⁸ s) timescale. Here we trace electronic dynamics with a time resolution of ≤ 150 as by using a subfemtosecond soft-X-ray pulse and a few-cycle visible light pulse. Our measurement indicates an attosecond response of the atomic system, a soft-X-ray pulse duration of 650 ± 150 as and an attosecond synchronism of the soft-X-ray pulse with the light field. The demonstrated experimental tools and techniques open the door to attosecond spectroscopy of bound electrons.

Experimental access to fast-evolving microscopic processes requires a short excitation ('pump') pulse, which sets the process going, and a short probe pulse for taking snapshots of the evolution of the process¹. Recent advances in ultrafast optics have made available laser pulses as short as a few femtoseconds for this time-resolved (or pump–probe) measurement technique^{2,3}. These pulse durations are practically at the limit set by the laser field oscillation cycle in the visible spectral range. They allow the tracking of changes in the nuclear structure of molecules—such as vibrations, or the breaking and formation of chemical bonds—because the characteristic timescale for atomic motion on an atomic length scale (0.1 nm) extends from a few femtoseconds to a few thousand femtoseconds (refs 4, 5). But time-domain access to a wide range of electron dynamics in the atomic shells has been frustrated so far by the speed of electronic relaxation processes. The wavefunction of bound electrons, following an energetic excitation—for example, ionization or the creation of an inner-shell vacancy—tends to evolve on an attosecond timescale. Until now, insight into these processes could be gained only indirectly from frequency-domain measurements of transition linewidths⁶.

Coherent extreme-ultraviolet (XUV) and X-ray sources are not only important tools for atomic spectroscopy but, owing to their short wave cycle, also offer the potential for producing electromagnetic radiation of attosecond duration. High-order harmonics of femtosecond laser radiation^{7,8} have been recently shown to be^{9,10} sources of trains of attosecond XUV pulses, separated by one half-cycle of the laser period with the train extending over some 10 fs. Yet, attosecond time-resolved measurements still appeared to be out of reach, because straightforward interpretation of spectroscopic data requires isolated, single pulses^{11,12}. Moreover, the XUV fluences available from current high-harmonic sources are many orders of magnitude lower than required by XUV-pump/XUV-probe spectroscopy, while conventional schemes of visible-light/XUV cross-correlation, yielding a convolution of the respective pulse envelopes, do not provide subfemtosecond time resolution. Using few-cycle visible laser pulses¹³ for the generation of isolated soft-X-ray pulses¹⁴, together with a special geometry for detecting photoelectrons originating from atoms exposed simultaneously to both of these pulses, our work offers a way to overcome these limitations.

Here we report the observation of light-field-induced petahertz frequency (1 PHz = 10¹⁵ Hz) modulation of the width of the kinetic energy distribution of photoelectrons; these photoelectrons were derived from an atomic gas (krypton, in our experiments) excited

by an ultrashort 90-eV soft-X-ray (henceforth briefly X-ray) pulse. Our measurements yield an X-ray pulse duration of $\tau_x = 650 \pm 150$ as, and indicate that more than 90% of the X-ray photons, which are spectrally filtered by a bandpass at ~90 eV, are confined within a single, isolated pulse of this duration. The experimental results bear direct evidence of the X-ray pulse being synchronized to the field oscillations of the visible light pulse with attosecond precision; they also indicate bound–free electronic transitions from the 4p state of krypton responding to 90-eV excitation on an attosecond timescale. With these tools we have traced the electric field oscillations in a visible light wave with a resolution of better than 150 as. These measurements constitute, to our knowledge, the first demonstrations of attosecond metrology. Beyond measuring the duration of X-ray pulses with attosecond resolution, our approach will permit attosecond spectroscopy of bound electron dynamics in atoms and molecules using a low-fluence subfemtosecond X-ray pulse and a strong visible light field.

X-ray photoemission controlled by a strong light field

Angle-resolved X-ray-induced photoemission assisted by visible light has been recently used to measure the cross-correlation of the intensity envelope of a 7-fs visible light pulse with that of an ultrashort soft-X-ray pulse¹⁴. Drawing on this previous work, we show here control of the final kinetic energy of X-ray-generated photoelectrons by the oscillating light field rather than by its cycle-averaged intensity. This enables us to probe, on an attosecond timescale, the duration of the X-ray pulse and its timing jitter with respect to the light field. The key to this progress has been (1) an approach that allows us to combine an enhanced solid angle of detection with attosecond temporal resolution, and (2) matching the X-ray pulse front to the light wave front within the volume of their interaction with nanometre-scale precision and stability. Photoelectrons from atoms exposed simultaneously to a strong light field and an X-ray pulse were detected within a cone aligned orthogonally to the beam direction and electric field vector of the linearly polarized light field. This orthogonal detection geometry was chosen to suppress a background in the photoelectron spectrum produced primarily along the electric field vector of the light pulse by above-threshold ionization (ATI).

The effect of the light field on the X-ray photoelectron energy spectrum is now well understood in a quantum-mechanical framework^{15,16}. In the strong-field limit, the most conspicuous features can also be accounted for by a quasi-classical model treating the interaction as a two-step process¹⁴: first, the photoelectron is

ejected by a short X-ray pulse with a distribution of initial momenta known from conventional photoionization studies⁶; and second, the photoelectron is subsequently accelerated as a classical particle by the electric field of a femtosecond light pulse. In accordance with the quantum treatment, the model predicts that, depending on the oscillation phase of the light field at the instant of 'birth' of the electron, a momentum component along the electric field vector, Δp_x , is added to the initial momentum of the electron, resulting in a shift of the photoelectron distribution up or down in momentum space (Fig. 1). The momentum transfer gives rise to a spread of the final kinetic energy of photoelectrons collected within a finite solid angle as revealed by the top and bottom of Fig. 1. The effect increases with increasing detection cone and vanishes for monodirectional detection.

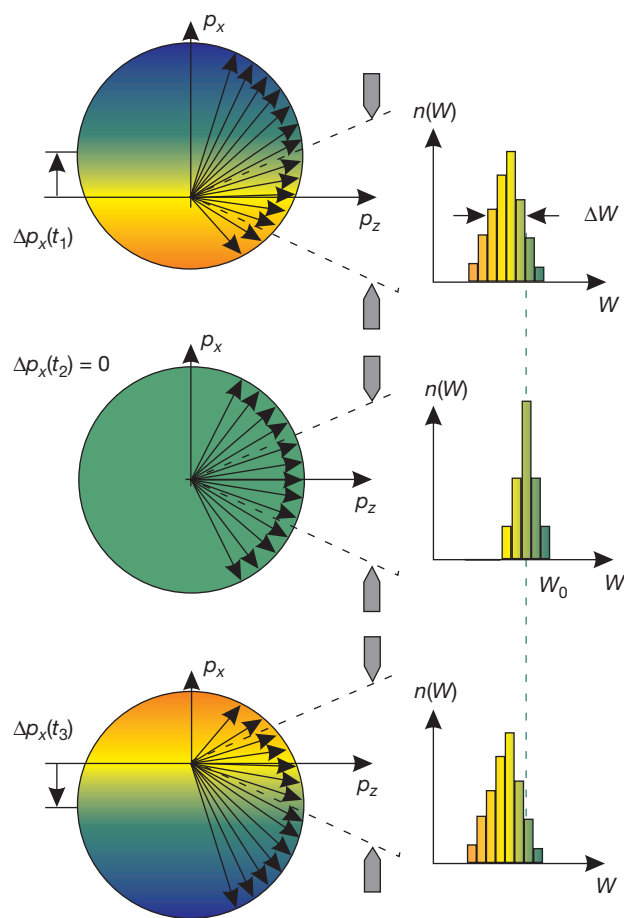


Figure 1 The principle of measuring cross-correlation between light electric field and X-ray intensity with attosecond resolution by means of 'two-colour' photoionization. The photoelectrons created initially with isotropic momentum distribution by the absorption of X-ray photons pick up momentum from the strong femtosecond laser field. In the slowly varying wave approximation, the field-induced momentum change Δp_x depends on the phase, amplitude and oscillation frequency of the light electric field at the instant of birth and is independent of the subsequent evolution of the light field. At birth instants t_1 , $t_2 = t_1 + T_0/4$ and $t_3 = t_1 + T_0/2$ chosen such that the light field crosses zero at t_1 and t_3 , the light-induced momentum change Δp_x deforms the final photoelectron momentum and energy distribution as shown. Different colours represent electrons of different final kinetic energy, with the energy increasing from the red to the blue. The momentum transfer and hence the change in the energy spectrum is largest if the electric field of the light pulse crosses zero at the moment of birth of the photoelectron (top and bottom). For a birth instant coinciding with the peak of the light electric field (centre) the final energy spectrum is unaffected by the light field and is determined by the spectrum of the X-ray pulse (represented by a single colour and a well-defined radius of the sphere in momentum space for simplicity).

Scanning the instant of birth of the photoelectron through the light field oscillations by changing the relative delay t_d between the light pulse and X-ray pulse results in a modulation of the spectral width $\Delta W(t_d)$ with a period equal to one-half of the light oscillation period T_0 . Figure 1 reveals that a confined detection cone aligned orthogonally to the light field vector also gives rise to a downshift of the centre of gravity of the photoelectron spectrum. This energy shift was used in our earlier study as a primary observable¹⁴. The key advantage of exploiting instead the broadening of the spectrum is that this effect is enhanced by increasing the detection aperture, and thus allows for a significantly enhanced signal yield and modulation amplitude. By contrast, the modulation in the centre of gravity of the photoelectron spectrum gets increasingly washed out with increasing detection angle, resulting in deterioration of the temporal resolution.

In the above description, we tacitly assumed an X-ray pulse duration τ_x that was very short compared to $T_0/2$, as well as an instantaneous response of the electronic transition, defining precisely the instant of birth of the photoelectron on a timescale of $T_0/2$. A finite X-ray pulse duration and timing jitter of the X-ray pulse (relative to the phase of the light field) of any origin results in a reduced depth of the above-discussed modulation of $\Delta W(t_d)$. In fact, an X-ray pulse duration approaching $T_0/2$ smears out the modulation completely. Our results rely on this line of argument: the modulation depth of $\Delta W(t_d)$ sets a reliable upper limit, on an attosecond timescale, on the X-ray pulse duration and timing jitter.

Cross-correlation of light electric field and X-ray intensity

The experiments reported below have been performed with a modified version of the apparatus described previously¹⁴. Major modifications included an enhanced detection angle and the installation of a feedback-controlled piezo delay stage as well as X-ray beam diagnostics. Figure 2 illustrates schematically the main steps in creating and processing the X-ray and light pulses for the photoemission experiment discussed in the preceding section.

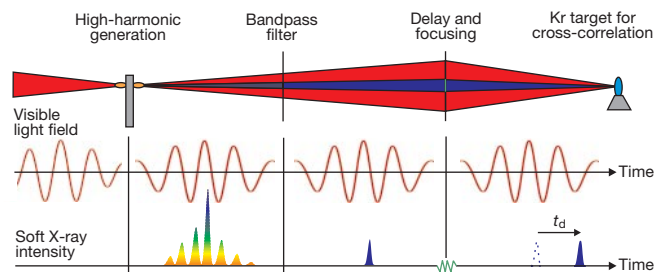


Figure 2 Diagram of the principal physical effects and processes preceding our light-field-controlled X-ray photoemission experiment. First, the few-cycle visible light pulse interacts with a jet of neon atoms to produce a train of pulses in the XUV and soft-X-ray spectral range. The rainbow colours of the individual pulses symbolize the respective spectral energy content: the highest photon energy components are contained only in a single pulse generated near the peak of the driver pulse. Note the increase of the instantaneous frequency in the centre of the visible light pulse (addressed as a dynamic blue shift in the text) past the interaction due to self phase modulation in the laser-generated plasma. The collinear laser and soft-X-ray beams are passed through a small circular Zr filter matched in size to the diameter of the harmonic beam. It stops the central portion of the laser beam, and transmits the low-divergence harmonic beam near 90 eV along with an annular laser beam surrounding it. The two co-axial beams then are focused onto the Kr atomic gas target by a concentric piezo-controlled double mirror unit¹⁴ (rays actually reflected at this point are unfolded in our schematic illustration for convenience). The central Mo/Si reflector serves simultaneously as a focusing mirror for the 90-eV soft-X-ray harmonics, and as a bandpass filter that selects the highest-energy photons contained only in the central harmonic pulse. Furthermore, this multifunctional unit provides the necessary adjustable delay between the visible light and soft-X-ray harmonic pulse.

The kinetic energy spectrum of $4p$ photoelectrons ejected from krypton atoms under simultaneous irradiation by the 90-eV X-ray pulse ($\lambda \approx 14$ nm) and the light pulse ($\lambda \approx 750$ nm) driving the harmonic radiation was measured as a function of t_d , which is the delay between the X-ray and the light pulse, both polarized along the x direction in Fig. 1. The Kr $4p$ electrons were chosen because of their favourable yield along the z direction at our X-ray photon energies⁶. The light pulse (peak intensity, $\sim 5 \times 10^{13}$ W cm⁻²) overlapped with the more tightly focused X-ray pulse in a volume of krypton gas kept at a pressure of a few times 10^{-2} mbar, limited to this value to prevent space-charge effects. The detection cone of our time-of-flight spectrometer is defined by the aperture of the spectrometer and its distance from the krypton target, with its full angle limited to $\sim 40^\circ$ by the numerical aperture of the electron lens system employed.

The delay was varied in steps of $\Delta t_d = 150$ as in the central range of temporal overlap (-5 fs $\leq t_d \leq 5$ fs, with $t_d = 0$ near the peak of the light pulse), and in larger steps outside this range. Figure 3 shows two representative Kr $4p$ spectra corrected for an ATI background and recorded at $t_d = -450$ as and $t_d = 0$, along with respective asymmetric gaussian fits. This model function yields excellent fits to our data, characterized by a fidelity parameter $R^2 > 0.95$. The noise increasing to lower energies can be attributed to ATI electrons. The contour plot in Fig. 4 depicts a series of Kr $4p$ photoelectron spectra (processed as described above) as a function of delay t_d . The data clearly bring to light a quasi-periodic evolution of the photoelectron energy spectrum. Both the width and, as a consequence of normalization to constant area, the peak height of the photoelectron energy spectrum are subject to a modulation versus t_d at a period of $\sim T_0/2$. Coupled to a modulation of the spectral width, we observe a periodic downshift of the centre of gravity of the spectrum, in agreement with our model's (Fig. 1) prediction for a narrow detection solid angle. However, the spectral shift is found to be contaminated by much larger uncertainties than is the variation of the spectral width.

For a quantitative analysis, we study the width of the photoelectron spectrum ΔW as a function of t_d . The upper panel of Fig. 4 shows the variation of the full-width at half-maximum $\Delta W(t_d)$ of the gaussian fits versus delay between the X-ray and light pulse. In what follows we focus on the oscillatory part of $\Delta W(t_d)$, which was obtained by subtracting the cycle-averaged 'd.c.' component of $\Delta W(t_d)$. The result is depicted by dots in Fig. 5 along with the respective error bars. From an inspection of the error bars at points of maximum slope in Fig. 5, we infer a time resolution of our

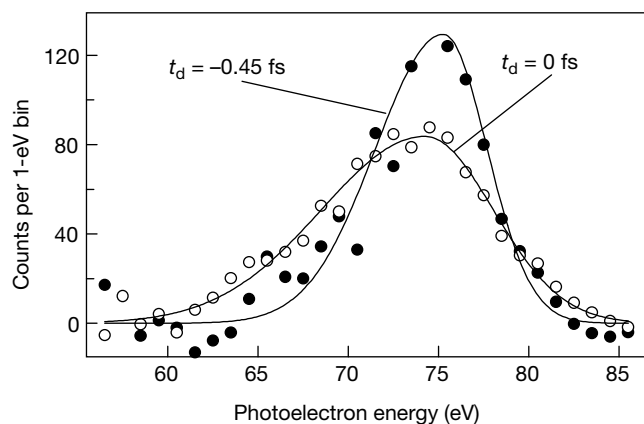


Figure 3 Kr $4p$ photoelectron spectra produced by ~ 90 -eV soft-X-ray pulses in the presence of a strong visible light field at two different delays t_d of the X-ray pulse. (Positive t_d corresponds to a retardation of the X-ray pulse.) The dots represent results corrected for ATI background. The lines show asymmetric gaussian fits to the data. The spectra were acquired over a measurement time of 2 minutes each.

sampling scheme of ≤ 150 as. The modulation period of the data is close to 1.25 fs in the wings of the correlation function $\Delta W(t_d)$, in agreement with a carrier wavelength of ~ 750 nm of the light pulse. However, it decreases to ~ 0.9 fs near zero delay, indicative of a pronounced blue shift of the light pulse at its peak, the origin of which we shall discuss below.

Subfemtosecond X-ray pulse

To evaluate the X-ray pulse duration from our light-field/X-ray-intensity cross-correlation data (dots in Fig. 5), we simulated the light-field-induced variation of the Kr $4p$ photoelectron spectrum versus t_d based on the quasi-classical model described above. The duration τ_x of our gaussian model X-ray pulse was used as the only fit parameter. In the analytic representation of the light pulse, the pronounced frequency upshift at the pulse centre and a small residual chirp apparent in the measured correlation function $\Delta W(t_d)$ was taken into account as described in the caption of Fig. 5. Under the assumption of an isotropic initial momentum distribution of X-ray photoelectrons and a uniform detection efficiency within the aperture of the spectrometer, our simple model reproduces well all observed key features (oscillating broadening and downshift of the photoelectron spectrum) except for a pronounced asymmetry of the broadened spectra.

The red line in Fig. 5 shows the results of our model calculations for an X-ray pulse length of $\tau_x = 650$ as, which—using no fit parameter other than τ_x —constitutes the best fit to the measured data in the range of maximum modulation depth (-1 fs $\leq t_d \leq 2$ fs). The inset displays, in addition, the computed correlation function $\Delta W(t_d)$ for $\tau_x = 500$ as and $\tau_x = 800$ as, revealing modulations respectively significantly enhanced and suppressed with respect to the measured data. These data provide a safe estimate for the duration of our 90-eV X-ray pulse within the range of $\tau_x = 650 \pm 150$ as, with the lower limit applying under the assumption of zero timing jitter of the X-ray pulse relative to the light wave.

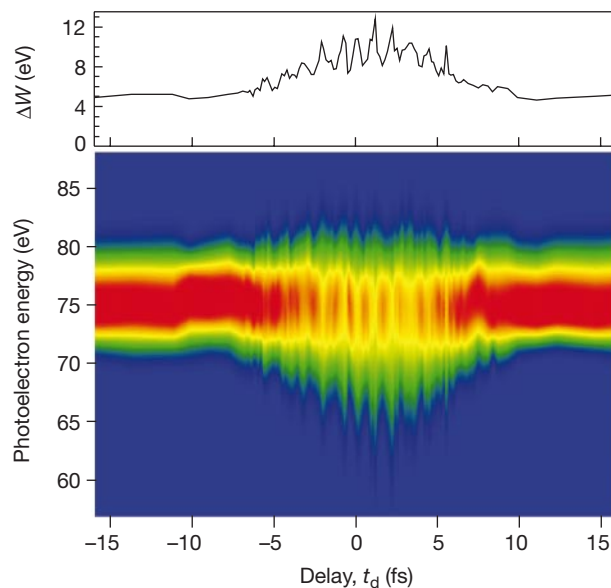


Figure 4 Cross-correlation of X-ray pulse with few-cycle laser pulse. Lower panel, contour plot of the Kr $4p$ photoelectron spectra (vertical axis) as a function of delay t_d of the soft-X-ray pulse with respect to the light pulse (horizontal axis). The plot displays the asymmetric gaussian fits introduced in Fig. 3, and comprises some 120 fitted spectra of the $4p$ feature, each normalized to the same number of counts, resulting in a constant area under the spectral distribution functions. Upper panel, spectral width ΔW evaluated from the gaussian fits as a function of the delay t_d . The spectra recorded at large delays, that is, those unaffected by the light field, are consistent with the X-ray spectrum centred at ~ 90 eV and with a binding energy of ~ 14 eV of the Kr $4p$ electron.

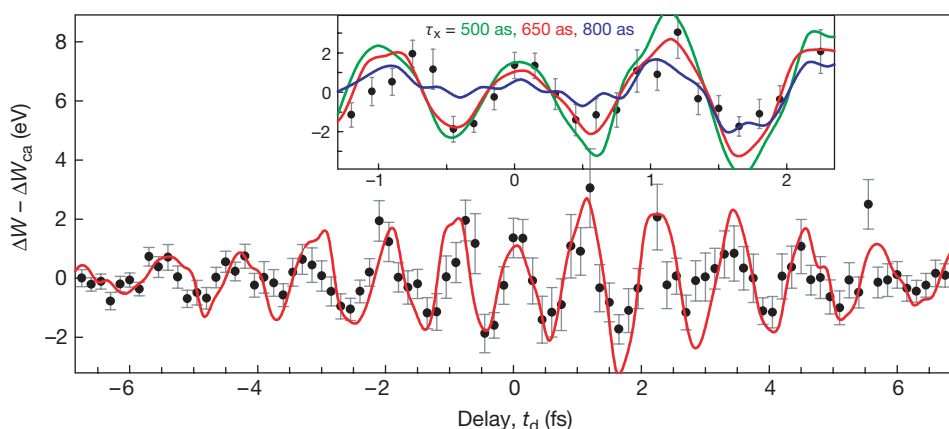


Figure 5 Spectral width ΔW of the Kr 4p photoelectron spectra as a function of t_d . The dots represent the oscillating component of the spectral width, $\Delta W - \Delta W_{ca}(t_d)$, where $\Delta W_{ca}(t_d)$ stands for the cycle-averaged value of $\Delta W(t_d)$. The full red line is the result of simulations based on the quasi-classical theory of two-colour X-ray photoionization by assuming a 7.5-fs, 750-nm linearly polarized light field of $5 \times 10^{13} \text{ W cm}^{-2}$ peak intensity and a 650-as, 90-eV gaussian X-ray pulse. The model light pulse includes a gaussian frequency sweep at the pulse peak superimposed on a weak quadratic component. Inset,

coloured lines show the result of the same simulation for different pulse durations near $t_d = 0$. The conspicuous asymmetry in the calculated and measured $\Delta W(t_d)$ about $t_d = 0$ is due to a change of the light carrier frequency occurring on a timescale T_0 . Consequently, the slowly varying wave approximation breaks down and the final state of the electron becomes dependent on the actual evolution of the light field. The substructure in the curves is an artefact, resulting from the statistical initial conditions in our simulations.

The emergence of a subfemtosecond pulse from high harmonic generation in the few-cycle regime was corroborated by numerical simulations of our X-ray source with a computer code¹⁷. This code solved Maxwell's wave equations in three dimensions and computed the radiation of the strongly-driven atomic dipoles using a generalization of the quantum treatment of ref. 18. These simulations yield a 530-as X-ray pulse near the propagation axis in the far field within a 5-eV spectral range near 90 eV (full line in Fig. 6), accompanied by a few small satellite pulses. The appearance of satellites separated by $\sim T_0/2$ from the central pulse leads to a spectral modulation with a period of twice the laser photon energy, as revealed by the calculated spectrum (full line) in the inset of Fig. 6. The depth of this modulation provides a sensitive measure of satellite content. The measured spectrum of the harmonic X-ray pulse reflected by our Mo/Si multilayer (dotted line in the inset of Fig. 6) sets a safe upper

limit of 10% to the total satellite fluence relative to the total X-ray fluence within a 5-eV range around 90 eV.

The agreement of the measured X-ray pulse duration with that obtained from our numerical calculations¹⁶ (within the experimental error) shows that the timing jitter of the X-ray pulse relative to the visible light field is much smaller than 1 fs. In other words, our experiment shows that the subfemtosecond X-ray pulse is locked to the carrier wave of its generating few-cycle light pulse with attosecond precision. This is surprising, given that the 'absolute' phase φ of the wave (defined in the caption of Fig. 6) with respect to the amplitude envelope is random in our laser pulses³, slightly modifying the evolution of the electric field from pulse to pulse. The attosecond timing stability of our subfemtosecond X-ray pulse to its few-cycle driver is due either to a (surprising) robustness of the high-harmonic generation process against these modifications, or (more likely) to a substantially reduced X-ray yield for values of φ other than the optimum assumed in Fig. 6. In the latter case, only a fraction of all laser pulses would provide an appreciable contribution to the detected photoelectron yield. If this is true, few-cycle driver pulses with stabilized absolute phase would markedly improve the data acquisition rate in the present and related experiments.

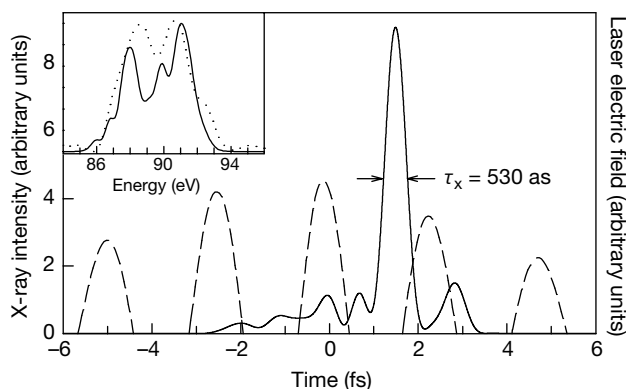


Figure 6 Calculated far-field, near-axis temporal intensity profile of a soft-X-ray pulse. The X-ray pulse is produced in a 3-mm-long 200-mbar neon gas volume by a 7-fs, 750-nm gaussian laser pulse with an on-axis peak intensity of $9 \times 10^{14} \text{ W cm}^{-2}$. For the electric field of the laser pulse, $E(t) \propto \exp(-t^2/\tau_L^2) \cos(\omega_0 t + \varphi)$ with $\varphi = 0$ (cosine pulse), where τ_L is the pulse duration, ω_0 is the angular carrier frequency and φ is the 'absolute' phase. The dashed line shows the on-axis electric field of the laser pulse leaving the interaction region. The calculated X-ray radiation is selected within a 5-eV spectral range near 90 eV. Inset, calculated (full line) and measured (dotted line) X-ray pulse spectrum selected by the Mo/Si reflector, showing that about 90% of the total fluence is within a 5-eV range around 90 eV.

Probing light-field oscillations

With its duration extracted from the central part of $\Delta W(t_d)$, the X-ray pulse can now be used, as a first application, to probe the evolution of the electric field in the light pulse. The sweep of the instantaneous carrier frequency ν_{inst} (or wavelength λ_{inst}) in the visible light pulse can be evaluated from $\Delta W(t_d)$ by fitting a sinusoidal oscillation of adjustable period to the data in Fig. 5. The dots in Fig. 7 show the carrier frequency sweep evaluated in this manner, revealing a dynamic blue shift from a carrier wavelength of $\sim 750 \text{ nm}$ to some 550 nm. To understand this phenomenon, we have to remember that the light pulse used here is derived from the one generating the X-ray pulse (Fig. 2). Also shown in Fig. 7 is the calculated frequency sweep of the laser pulse leaving the neon-gas X-ray source.

The measured dynamic frequency shift at the centre of the pulse is larger than predicted by our numerical study, but reflects qualitatively the predicted behaviour. The qualitative agreement suggests that the observed blue shift can be attributed to ionization-induced

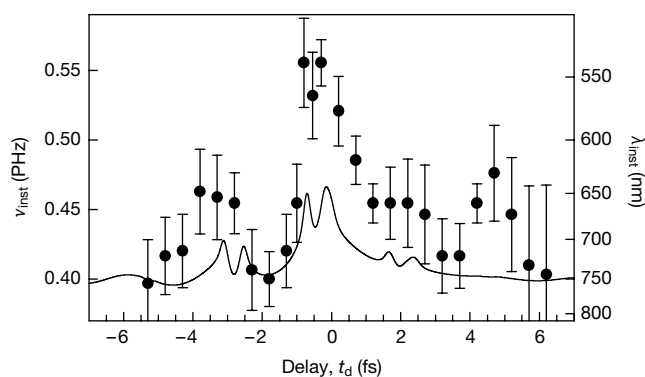


Figure 7 Calculated (line) and measured (dots) instantaneous frequency of the few-cycle light field. The experimental data are derived from $\Delta W(t_d)$ shown in Fig. 5. The computed frequency sweep originates from self-phase modulation of the light pulse upon its interaction with the neon gas X-ray source.

self-phase modulation in the X-ray generation process. These results provide evidence of optical field ionization driven by a few-cycle pulse being confined to less than one single oscillation cycle. The direct probing of the field oscillations in a light wave as implemented here is another demonstration of attosecond metrology. It should permit complete measurement of the electric field of absolute-phase-stabilized few-cycle light^{19,20}. The presence of a substantial dynamic blue shift in the light pulse used to control X-ray photoemission provides a further strong argument for the isolated nature of our X-ray pulse: the depth of modulation of $\Delta W(t_d)$ near $t_d = 0$, where the modulation period drops to ~ 0.9 fs, would rapidly decrease with increasing satellites spaced by ~ 1.25 fs from the main pulse (Fig. 6).

Attosecond spectroscopy

With the availability of single subfemtosecond X-ray pulses, it is now possible to extend time-resolved spectroscopy into the attosecond domain by using these pulses for both triggering and probing bound-bound or bound-free electronic transitions in atoms or molecules. However, the subfemtosecond X-ray pulses currently available do not yet have sufficient fluence for X-ray-pump/X-ray-probe spectroscopy. Generalization of the present concept of light-field-controlled X-ray photoemission should allow the substitution of either the X-ray pump or the X-ray probe pulse by a strong few-cycle laser field without losing attosecond resolution. This approach vastly relaxes requirements on the X-ray pulse fluence, holding promise for extending the technique up to photon energies approaching one kiloelectron volt (keV) with next-generation high-harmonic sources.

In one class of proposed experiments, an absolute-phase-stabilized¹⁹ quasi-single-cycle visible laser pulse (for example, $\tau_p = 4$ fs at $\lambda = 750$ nm, as shown by Fig. 20 in ref. 3) sets free a subfemtosecond wavepacket by optical field ionization; this wavepacket is driven back to its parent ion an optical cycle later with energies of up to a few keV and produce collisional excitations. The relaxation processes following subfemtosecond collisional excitation are important for a number of phenomena, such as X-ray lasing

and non-sequential photoionization. These processes could be traced in time by measuring the energy of photoelectrons detached from the ion by a delayed, subfemtosecond X-ray probe pulse. In another class of studies, the subfemtosecond X-ray pulse might serve as a pump pulse to create inner-shell vacancies. The temporal evolution of the subsequent inner-shell relaxation processes could be traced by time-resolved spectroscopy of the emitted Auger electron(s). The Auger spectrum, modified by a few-cycle light field and measured as a function of t_d , would provide access to the (few-femtosecond to attosecond) inner-shell relaxation times in the range of several hundred eV to keV in the same way as the correlation functions in Figs 4 and 5 yield the X-ray pulse duration. The experimental tools are now becoming available for attosecond spectroscopy. □

Received 1 October; accepted 19 October 2001.

- Krausz, F. From femtochemistry to attophysics. *Phys. World* **14**(9), 41–46 (2001).
- Steinmeyer, G., Sutter, D. H., Gallmann, L., Matuschek, N. & Keller, U. Frontiers in ultrashort pulse generation: pushing the limits in linear and nonlinear optics. *Science* **286**, 1507–1512 (1999).
- Brabec, T. & Krausz, F. Intense few-cycle laser fields: frontiers of nonlinear optics. *Rev. Mod. Phys.* **72**, 545–591 (2000).
- Zewail, A. Femtochemistry: atomic-scale dynamics of the chemical bond (adapted from the Nobel Lecture). *J. Phys. Chem. A* **104**, 5660–5694 (2000).
- Bhattacharjee, Y. Measuring the immeasurable. *Nature* **412**, 474–476 (2001).
- Becker, U. & Shirley, D. A. *VUV and Soft X-Ray Photoionization* 152 (Plenum, New York, 1996).
- L'Huillier, A. & Balcou, P. High-order harmonic generation in rare gases with a 1-ps 1053-nm laser. *Phys. Rev. Lett.* **70**, 774–777 (1993).
- Macklin, J. J., Kmetec, J. D. & Gordon, C. L. High-order harmonic generation using intense femtosecond pulses. *Phys. Rev. Lett.* **70**, 766–769 (1993).
- Papadogiannis, N. A., Witzel, B., Kalpouzos, C. & Charalambidis, D. Observation of attosecond light localization in higher order harmonic generation. *Phys. Rev. Lett.* **83**, 4289–4292 (1999).
- Paul, P. *et al.* Observation of a train of attosecond pulses from high harmonic generation. *Science* **292**, 1689–1692 (2001).
- Ivanov, M., Corkum, P. B., Zuo, T. & Bandrauk, A. Routes to control of intense-field atomic polarizability. *Phys. Rev. Lett.* **74**, 2933–2936 (1995).
- Christov, I. P., Murnane, M. M. & Kapteyn, H. C. High-harmonic generation of attosecond pulses in the “single-cycle” regime. *Phys. Rev. Lett.* **78**, 1251–1254 (1997).
- Spielmann, Ch. *et al.* Generation of coherent X-rays in the water window using 5-femtosecond laser pulses. *Science* **278**, 661–664 (1997).
- Drescher, M. *et al.* X-ray pulses approaching the attosecond frontier. *Science* **291**, 1923–1927 (2001). Published online 15 February 2001; 10.1126/science.1058561.
- Kitzler, M., Milosevic, N., Scrinzi, A. & Brabec, T. Theory of attosecond streak cameras. *Phys. Rev. Lett.* (submitted).
- Itatani, J. *et al.* Attosecond streak camera. *Phys. Rev. Lett.* (submitted).
- Milosevic, N., Scrinzi, A. & Brabec, T. Ab initio numerical calculation of attosecond pulse generation. *Phys. Rev. Lett.* (submitted).
- Lewenstein, M., Balcou, Ph., Ivanov, M. Yu., L'Huillier, A. & Corkum, P. B. Theory of high-harmonic generation by low-frequency laser fields. *Phys. Rev. A* **49**, 2117–2132 (1994).
- Poppe, A. *et al.* Few-cycle optical waveform synthesis. *Appl. Phys. B* **72**, 373–376 (2001). Published online 13 December 2000; 10.1007/s003400000526.
- Paulus, G. G. *et al.* Evidence of “absolute-phase” phenomena in photoionization with few cycle laser pulses. *Nature* **414**, 182–184 (2001).

Acknowledgements

We thank M. Uiberacker for assistance with measurements, and Y. Lim and U. Kleineberg for manufacturing the X-ray multilayer mirror. Discussions with M. Ivanov are gratefully acknowledged. This work was supported by the Austrian Science Fund and by the European ATTO network.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.K. (e-mail: ferenc.krausz@tuwien.ac.at).

Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication

Piers Nash^{*†}, Xiaojing Tang^{*†}, Stephen Orlicky^{*}, Qinghua Chen[‡], Frank B. Gertler[§], Michael D. Mendenhall[‡], Frank Sicheri^{*||}, Tony Pawson^{*||} & Mike Tyers^{*||}

^{*} Programme in Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto M5G 1X5, Canada

[‡] L.P. Markey Cancer Center, Department of Biochemistry, University of Kentucky, Lexington, Kentucky 40536-0096, USA

[§] Biology Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

^{||} Department of Medical Genetics and Microbiology, University of Toronto, 1 Kings College Circle, Toronto M5S 1A8, Canada

[†] These authors contributed equally to this work

SCF ubiquitin ligases target phosphorylated substrates for ubiquitin-dependent proteolysis by means of adapter subunits called F-box proteins. The F-box protein Cdc4 captures phosphorylated forms of the cyclin-dependent kinase inhibitor Sic1 for ubiquitination in late G1 phase, an event necessary for the onset of DNA replication. The WD40 repeat domain of Cdc4 binds with high affinity to a consensus phosphopeptide motif (the Cdc4 phospho-degron, CPD), yet Sic1 itself has many sub-optimal CPD motifs that act in concert to mediate Cdc4 binding. The weak CPD sites in Sic1 establish a phosphorylation threshold that delays degradation *in vivo*, and thereby establishes a minimal G1 phase period needed to ensure proper DNA replication. Multisite phosphorylation may be a more general mechanism to set thresholds in regulated protein–protein interactions.

Numerous regulatory proteins are targeted for degradation in a precisely programmed manner through the covalent conjugation of ubiquitin, which is transferred along a cascade of E1, E2 and E3 enzymes to the substrate¹. Reiterative transfer of ubiquitin generates polyubiquitinated species that are recognized and rapidly degraded by the 26S proteasome. E3 enzymes, or ubiquitin ligases, catalyse the terminal step in ubiquitin transfer, and as such are the crucial determinants of substrate specificity. Substrate recognition depends on often ill-defined sequence elements, referred to as degrons, that are the binding sites for cognate E3 enzymes^{1,2}. The E3–substrate interaction can be regulated at several levels. In some instances, limiting cofactors determine E3 activity, as in the case of the anaphase promoting complex/cyclosome (APC/C), which targets mitotic cyclins and other proteins for degradation during mitosis³. In other cases, E3 recognition depends on post-translational modification of the substrate. In particular, phosphorylation is often used to direct substrates to a recently described class of E3 enzymes termed Skp1–Cdc53/Cul1–F-box protein (SCF) complexes³. SCF complexes target a broad spectrum of substrates through a repertoire of substrate-specific adapter subunits called F-box proteins⁴. The 40-amino-acid F-box motif is a binding site for Skp1, which in turn links F-box proteins to a core ubiquitination complex composed of the scaffold protein Cdc53/Cul1, the RING-H2 domain protein Rbx1 (also known as Roc1 or Hrt1) and, usually, the E2 enzyme Cdc34 (ref. 3). F-box proteins capture phosphorylated substrates by means of carboxy-terminal protein–protein interaction regions, such as WD40 repeat domains or leucine-rich repeat (LRR) domains⁵.

Cell cycle progression requires the precisely ordered elimination of cyclins and cyclin-dependent kinase (CDK) inhibitors by the ubiquitin system³. In yeast, commitment to division, called Start, requires a threshold level of G1 cyclins (Cln1/2/3), which serve to activate Cdc28 (also called Cdk1) in late G1 phase. As cells pass Start, B-type cyclin (Clb5/6)–Cdc28 kinases are activated, which is a necessary step for initiation of DNA replication⁶. The primary function of Cln–Cdc28 activity is to phosphorylate an inhibitor of the Clb–Cdc28 kinases called Sic1, thereby targeting it for

degradation^{6,7}. Phospho-Sic1 is specifically recognized by the F-box protein Cdc4, which recruits Sic1 for ubiquitination by the Cdc34–SCF complex^{4,5,8}. Overexpression of stabilized forms of Sic1 that lack Cdc28 phosphorylation sites cause an arrest at the G1 phase⁹, whereas deletion of *SIC1* causes premature DNA replication and rampant genome instability¹⁰. Cdc4 recruits several other substrates to the SCF core complex in a phosphorylation-dependent manner, including the Cln–Cdc28 inhibitor/cytoskeletal scaffold protein Far1, the replication protein Cdc6 and the transcription factor Gcn4 (ref. 3). In the mammalian cell cycle, SCF complexes target phosphorylated forms of cyclin E1 and the CDK inhibitor p27^{Kip1} (refs 11, 12). The important role of SCF pathways is shown by the G1 phase arrest caused by non-phosphorylatable forms of p27^{Kip1}, and by the genome instability caused by expression of stabilized forms of cyclin E1 (refs 13, 14). In addition to cell cycle control, SCF-dependent proteolysis regulates the NFκB and Wnt/β-catenin signalling pathways, among others¹⁵.

Despite the well documented requirement for substrate level phosphorylation in SCF-dependent ubiquitination, the mechanism by which phosphorylation drives substrate binding is unclear. On the one hand, for some substrates the binding interaction seems to be based on recognition of a phosphopeptide motif^{16–18}, in a manner that is analogous to phosphorylation-dependent recognition by Src homology 2 (SH2), phosphotyrosine-binding (PTB), 14-3-3 and forkhead-associated (FHA) domains¹⁹. On the other hand, degradation of substrates such as Sic1, Cdc6 or Cln2 seems to require phosphorylation on multiple, widely spaced sites^{9,20,21}. Notably, alignment of the numerous genetically relevant phosphorylation sites in such substrates does not reveal an obvious consensus binding motif. To address the mechanism of Cdc4 substrate recognition, we surveyed a number of synthetic phosphopeptides derived from known substrates and identified a high-affinity consensus binding motif, termed the Cdc4 phospho-degron (CPD). The endogenous Cdc4 substrate Sic1 contains nine suboptimal CPD sites, at least six of which must be phosphorylated to allow recognition by Cdc4. This requirement for multisite phosphorylation not only imposes a threshold for Cln–Cdc28 kinase in late G1

phase, but may also confer switch-like characteristics on Sic1 degradation and subsequent entry into S phase.

Phosphorylation requirements in Sic1 recognition

Previous analysis has shown that multiple phosphorylation sites contribute to Sic1 ubiquitination *in vitro* and degradation *in vivo*⁹. To systematically determine the relative contributions of each of the nine CDK consensus sites in Sic1, we mutated each individual site and assessed the mutant proteins for stability *in vivo*, as well as for binding to Cdc4 *in vitro* (Fig. 1). Each of the mutant proteins inhibited Clb5–Cdc28 kinase activity to the same extent as wild-type Sic1, including Sic1^{0p}, which lacks all nine phosphorylation sites (Fig. 1b). Consistent with previous studies^{9,22}, conditional overexpression of *SIC1*^{T45A} from the *GAL1* promoter permanently arrested cells in G1 phase, whereas expression from the wild-type promoter caused a modest G1 delay (data not shown). Repression of the various *GAL1-SIC1* constructs in cells arrested at Start by mating pheromone allowed an estimate of Sic1 half-life *in vivo*, using a curve-fitting algorithm (Fig. 1c). Several phosphorylation sites contributed to Sic1 instability, with a rank order of Thr 45, Ser 76, Thr 5 and Thr 33, followed by less significant contributions from other sites. Analysis of binding of the panel of Sic1 mutants to

Cdc4 indicated that no single site is required for recognition, although loss of either the Thr 45 or Ser 76 sites did diminish *in vitro* binding to some extent (Fig. 1d). We then determined which individual sites, or combinations thereof, are sufficient for degradation. Beginning with the Sic1^{0p} mutant, which lacks all nine CDK sites, we restored increasing numbers of phosphorylation sites and assessed the effects on viability and the Cdc4–Sic1 interaction. Serial reintroduction of the top-ranked five sites failed to restore Sic1 binding to Cdc4 or degradation *in vivo* (Fig. 1e, f). However, introduction of either of two combinations of six CDK phosphorylation sites or seven sites proved sufficient for Cdc4 binding *in vitro* and Sic1 degradation *in vivo* (Fig. 1e, f). Thus, efficient phosphorylation-dependent recognition by Cdc4 minimally requires six of the nine CDK phosphorylation sites in Sic1.

Identification of a Cdc4 phospho-degron

There are at least three possible modes of phospho-Sic1 binding to Cdc4. The first is a phosphorylation-dependent conformational change that exposes a pre-existing cryptic binding epitope on Sic1. The second is direct binding of multiple phosphorylated residues to multiple, distinct binding sites on Cdc4. The third mode is equilibrium binding of multiple phosphorylated residues on Sic1 with a single phospho-recognition site on Cdc4. To investigate these three different possibilities, we examined the ability of various synthetic phosphopeptides to bind to Cdc4 *in vitro* by fluorescence polarization. Phosphopeptides corresponding to sequences centred on Thr 45 of Sic1 or another candidate interaction site in Far1 (ref. 23) could neither bind to Cdc4 in this assay, nor block the interaction between full-length, phosphorylated substrates and Cdc4 (Fig. 2a, Table 1 and data not shown). We then surveyed other phosphopeptides and discovered that a 19-residue peptide centred on Thr 380 of mammalian cyclin E1 (CycE^{19pT380}), a site previously implicated in cyclin E1 degradation^{24,25}, bound to Cdc4 with high affinity. A biotinylated version of CycE^{19pT380}, but not the corresponding unphosphorylated peptide, was able to capture Cdc4 from lysates of insect cells infected with a recombinant baculovirus (see Supplementary Information). Fluorescence polarization measurements performed with CycE^{19pT380} and purified recombinant Cdc4 revealed an equilibrium dissociation constant (K_d) of $1.0 \pm 0.05 \mu\text{M}$ and a Hill coefficient of 0.99 for the interaction, indicating a single class of high-affinity binding site on Cdc4 (Fig. 2b). The phosphorylated Thr 380 site in cyclin E1 also mediates Cdc4 recognition within the context of the intact protein, as full-length cyclin E1 is targeted for degradation by SCF^{Cdc4}, both *in vitro* and *in vivo* (see Supplementary Information). Importantly, the CycE^{19pT380} phosphopeptide was able to out-compete both Cdc4 binding and ubiquitination of cyclin E1, Sic1 and Far1 (Fig. 2a and data not shown). Thus, the CycE^{19pT380} peptide binds the same site(s) on Cdc4 as its fully phosphorylated, physiological substrates.

The CPD consensus

To identify the principal phosphopeptide determinants for Cdc4 recognition, a peptide Spots array technique was used²⁶. Each position of the CycE^{19pT380} peptide was systematically altered to each of the 20 natural amino acids in a membrane-array format. Interaction of purified Skp1–Cdc4 complex with peptides on the membrane was detected with an anti-Skp1 antibody (Fig. 2c). Several characteristics of the binding site were revealed by the peptide array analysis. First, phosphorylated threonine (pThr) and a Pro at the +1 position are strictly required, consistent with the specificity of the targeting CDK kinases. Second, binding specificity is contributed by sequences that are amino-terminal to the phosphorylation site as there is a strong selection for Leu, Ile or Pro at the –1 position, whereas only Leu or Ile are preferred at the –2 position. Third, and quite unexpectedly, Arg and Lys residues seem to be disfavoured at the +2 to +5 positions, as is Tyr to a lesser extent. The optimal substrate selectivity of Cdc4 is therefore at odds

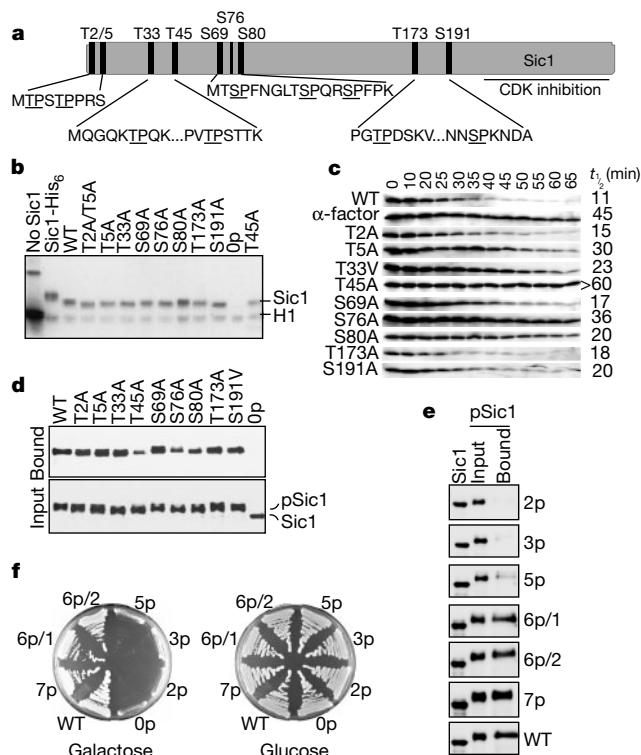


Figure 1 Contribution of CDK phosphorylation sites to Sic1 recognition, ubiquitination and degradation. **a**, Consensus S/T-P CDK phosphorylation sites in Sic1 (underlined). **b**, Inhibition of Clb5–Cdc28 kinase activity against histone H1 by purified Sic1 phosphorylation site mutants. WT, wild type. **c**, Half-life of Sic1 phosphorylation site mutants determined by decay of Sic1 protein on repression of *GAL1-SIC1*^{T45A} constructs after release from α -factor arrest. The row labelled α -factor indicates decay rate of wild-type Sic1 before release. **d**, Binding of Sic1 phosphorylation site mutants to purified Cdc4. GST–Sic1 fusion proteins were phosphorylated by recombinant Cln2–Cdc28, captured onto Flag-tagged Cdc4 resin and detected with anti-Sic1 antibody. **e**, **f**, Phosphorylation of a minimum of six sites on Sic1 is required for interaction with Cdc4 *in vitro* (**e**) and degradation of Sic1 *in vivo* (**f**). Sites are indicated as follows: 2p = Thr 45, Ser 76; 3p = Thr 33, Thr 45, Ser 76; 5p = Thr 2, Thr 5, Thr 33, Thr 45, Ser 76; 6p/1 = Thr 2, Thr 5, Thr 33, Thr 45, Ser 69, Ser 76; 6p/2 = Thr 2, Thr 5, Thr 33, Thr 45, Ser 76, Ser 80; 7p = Thr 2, Thr 5, Thr 33, Thr 45, Ser 69, Ser 76, Ser 80. *GAL1-SIC1* strains were incubated for 2 days at 30 °C.

with that of the cognate targeting kinase Cdc28, which strongly prefers to phosphorylate S/T-P sequences followed by C-terminal basic residues²⁷.

The peptide Spots analysis was verified by quantitative fluorescence polarization measurements with various derivatives of the CycE^{19pT380} phosphopeptide (Table 1). The minimal peptide sequence required for binding was delimited to a core recognition sequence, LLpTPP, which bound Cdc4 with a K_d of $0.85 \pm 0.1 \mu\text{M}$. Introduction of basic residues in the +2 to +5 positions caused a decrease in solution binding affinity, most notably at the +2 and +3 positions, in conformity to the sequence preference for phosphorylation directed by CDK enzymes. The detrimental effect of basic residues is mediated by positive electrostatic potential because placement of an acetylated lysine at the +3 position yielded wild-type binding affinity. Finally, the essential pThr residue was strongly preferred over pSer (Table 1), suggesting an additional level of complexity in substrate discrimination. For brevity, we refer to the consensus binding sequence, L/I-L/I/P-P-T-P(RK)₄, as the CPD motif, where (X) refers to disfavoured residues.

Physiological CPD motifs bind over a wide range of affinities

Database searches revealed that the CPD motif is present in

numerous yeast proteins, including a recently characterized SCF^{Cdc4} substrate—the yeast transcriptional activator Gcn4 (ref. 28). Notably, one of the relevant targeting phosphorylation sites on Gcn4, Thr 165, is embedded in a sequence that closely matches the CPD consensus, and a phosphopeptide centred on Thr 165 of Gcn4 bound to Cdc4 with a K_d of $0.88 \pm 0.1 \mu\text{M}$ (Table 1). When the Gcn4 peptide sequence was permuted in a Spots array, the resulting consensus binding sequence closely matched that derived from cyclin E1 (Fig. 2d). The Gcn4 sequence seems to be exquisitely tuned for Cdc4 binding, as variations in the principal CPD residues at positions -2 and -1 are not tolerated as well as they are in the core cyclin E1 motif. By the same token, residues at the +2 to +5 positions appear less important, presumably because the core region binds more tightly. Another known Cdc4 substrate, Far1, also contained two reasonable matches to the CPD motif, but not within regions previously implicated in Far1 degradation^{23,29}. Sequences centred on Thr 63 and Thr 306 matched the CPD motif, and indeed a phosphopeptide centred on Thr 306 bound weakly to Cdc4 (Table 1). Phosphorylation of this site seems to contribute to activation of Far1 by the MAP kinase Fus3 (ref. 30), raising the possibility that Far1 activation is directly coupled to its recognition by SCF^{Cdc4}. To directly compare the relative affinities of

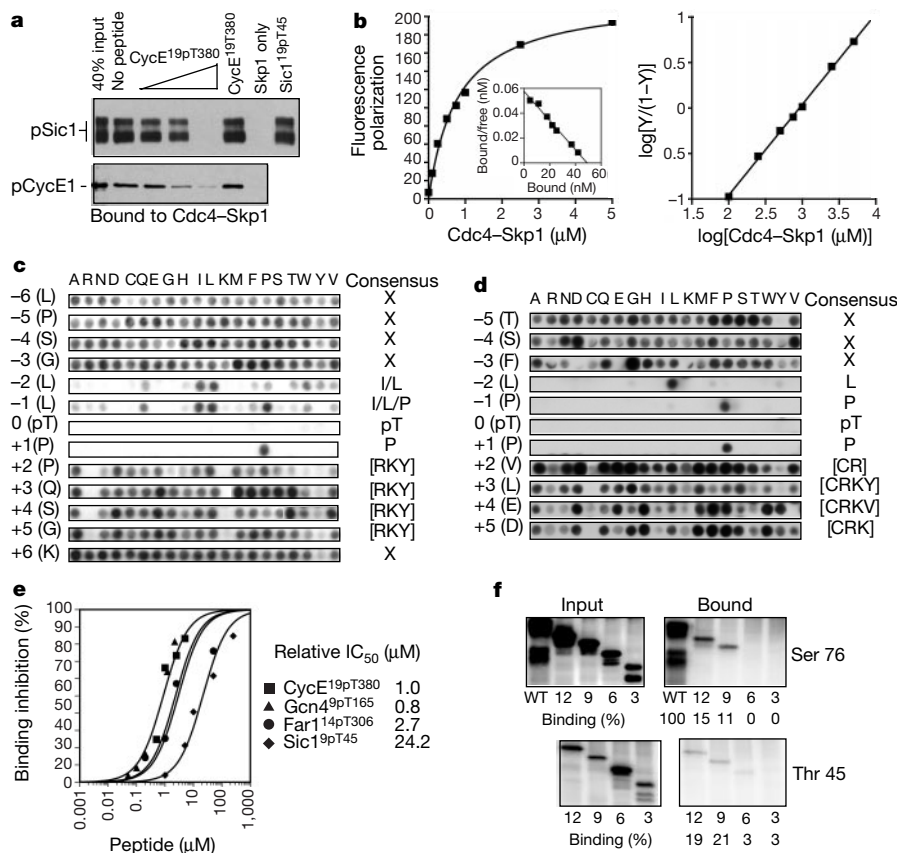


Figure 2 A phosphopeptide derived from cyclin E1 defines a high-affinity binding site on Cdc4. **a**, A CycE^{19pT380} phosphopeptide (ASPLPSGLpTPPQSGKKQS), but not a Sic1^{19pT45} phosphopeptide, out-competes binding of phospho-Sic1 and phospho-cyclin E1 to Cdc4. The indicated peptides (3, 17 and 68 μM for CycE^{19pT380}; 68 μM for CycE^{19pT380} and Sic1^{19pT45}) and phosphoproteins were incubated with Flag-tagged Skp1-Cdc4 resin. Bound proteins were detected with anti-Sic1 and anti-cyclin E1 antibodies. **b**, Michaelis-Menton plot (left), Scatchard plot (inset) and Hill plot (right) for the CycE^{19pT380} phosphopeptide interaction with Skp1-Cdc4 as measured by fluorescence polarization. **c**, **d**, Delineation of the CPD consensus with membrane-bound arrays of synthetic peptides based on the CycE^{19pT380} sequence (**c**) or the Gcn4^{11pT165} sequence (**d**). Each residue in the seed sequence (vertical axis) was substituted with every natural amino acid

(horizontal axis). Membranes were incubated with purified Skp1-Cdc4 complex followed by detection with anti-Skp1 antibody. **e**, Peptide half-maximal inhibitory concentration (IC₅₀) curves for competition of fluorescein-CycE^{9pT380} peptide away from Cdc4 by CycE^{19pT380}, Gcn4^{9pT165}, Far1^{14pT306} or Sic1^{19pT45}. **f**, Concatamerization of low-affinity CPD motifs generates a high-affinity interaction with Cdc4. Varying numbers of suboptimal CPD sites corresponding to sequences derived from either a mixed Sic1 S76/S80/T5S site (GLTSPQRSPFPKSSPPRS) or the Sic1 T45 site (VTPSKPVTSPKVTSPSR) were produced as GST fusion proteins, phosphorylated by Cln2/Cdc28 in the presence of [³²P]-γ-ATP, captured onto Flag-tagged Cdc4 resin and detected by autoradiography. Capture efficiencies were normalized to that of phospho-Sic1.

several CPD-containing peptides, we used a competition binding assay with a fluorescein-labelled CycE^{9pT380} peptide. Under these conditions, it was possible to detect weak binding of a Sic1^{9pT45} peptide, demonstrating that weak CPD sites can mediate a direct interaction with Cdc4. Other natural CPD sequences bound with various affinities in the rank order Gcn4^{9pT165} ≥ cyclin E^{19pT380} > Far1^{21pT306} >> Sic1^{9pT45} (Fig. 2e).

Concatamers of weak CPDs bind Cdc4 with high affinity

To test the possibility that the multiple, weak phospho-dependent interaction motifs in Sic1 might act in concert to form a high-affinity phospho-Sic1–Cdc4 complex, we tested artificial substrates that carried increasing numbers of weak CPD sites derived from Sic1. Glutathione S-transferase (GST) fusion proteins were constructed with 3, 6, 9 or 12 CDK phosphorylation sites, with the reiterated sequences corresponding to either an artificial fusion of the Ser 76, Ser 80 and Thr 5 (substituted to Ser) sites of Sic1, or to a variant form of the Thr 45 site with a disfavoured Lys residue substituted at the +3 position. The purified fusion proteins were phosphorylated with Cln2–Cdc28 kinase in the presence of [³²P]-γ-ATP and assayed for Cdc4 binding. Although fusion proteins tagged with three or six weak CPD sites proved incapable of interacting with Cdc4, those with 9 or 12 sites were efficiently captured by Cdc4 (Fig. 2f). All non-phosphorylated versions of the GST–CPD fusions failed to interact with Cdc4 (data not shown). It is unlikely that the two different polymers of weak CPD sites fold into a defined tertiary structure; indeed, circular dichroism studies suggest that in its unphosphorylated form Sic1 lacks any detectable higher-order structure (see Supplementary Information). On the basis of these results, we conclude that the Sic1–Cdc4 interaction is directly mediated by a series of low-affinity CPD sites.

CPD interactions with the WD40 domain

Deletion analysis of Cdc4 demonstrated that the WD40 repeat domain is sufficient for high-affinity interaction with the CycE^{19pT380} peptide (Fig. 3a). In all known cases, direct phospho-dependent interactions are mediated by Arg and Lys residues³¹. To

identify potential CPD binding sites, we modelled the Cdc4 WD40 repeat sequence onto the β-transducin WD40 structure³², and mapped eight Arg residues conserved between Cdc4 and its homologues from *Candida albicans* and *Schizosaccharomyces pombe* onto the predicted β-propeller structure (Fig. 3b and Supplementary Information). Each Arg residue was individually mutated to Ala, and the resulting mutants were tested for their ability to rescue Cdc4 function in a plasmid shuffle experiment. Three substitutions abrogated Cdc4 function *in vivo* (R467A, R485A and R534A), whereas the other five mutations did not alter growth rate (Fig. 3c). Recombinant R467A, R485A and R534A mutant proteins were severely compromised for binding to both the CycE^{19pT380} peptide and phosphorylated Sic1 protein, whereas the neutral mutation, R443A, did not alter binding (Fig. 3d, e). Thus, consistent with the single class of high-affinity binding site identified by fluorescence polarization, we have found a single, localized interaction region in Cdc4. We speculate that the triad of essential Arg residues may comprise a phospho-Thr binding pocket.

A single, optimal CPD efficiently targets Sic1

To test whether an optimal CPD can function in a heterologous context, we inserted the cyclin E1 peptide motif or derivatives thereof into the Sic1^{0p} mutant. When the CycE^{19pT380} sequence was placed at the Thr 45 site of Sic1^{0p} (Sic1^{T45::CycE19T380}), it was able to confer recognition and ubiquitination by SCF^{Cdc4} *in vitro* (Fig. 4a, b), and elimination of Sic1^{0p} *in vivo* (Fig. 4c). Mutation of the

Table 1 Affinities of phosphopeptides for Cdc4

Peptide name	Peptide sequence	K _d (μM)
Effect of peptide length		
CycE ^{19pT380}	ASPLPSGLLP TPPQSGKKQS	1.0 ± 0.08
CycE ^{16pT380}	ASPLPSGLLP TPPQSGK	0.9 ± 0.1
CycE ^{9pT380}	GLLP TPPQSG	1.0 ± 0.05
CycE ^{5pT380}	LLP TPP	0.85 ± 0.1
Effect of pT		
CycE ^{9pT380}	GLLP TPPQSG	1.0 ± 0.05
CycE ^{9pT380S}	GLLP SPPQSG	6.0 ± 0.9
CycE ^{9pT380Y}	GLLP YPPQSG	ND
CycE ^{9pT380}	GLLP TPPQSG	ND
Cost of basic residues		
CycE ^{9pT380}	GLLP TPPQSG	1.0 ± 0.05
CycE ^{9pT380/L378K}	GKLP TPPQSG	12 ± 2
CycE ^{9pT380/L379K}	GLKLP TPPQSG	6.0 ± 1.2
CycE ^{9pT380/P381A}	GLLP TAPQSG	ND
CycE ^{9pT380/P381K}	GLLP TPKQSG	6.3 ± 1.2
CycE ^{9pT380/Q382K}	GLLP TPKQSG	5.1 ± 1.4
CycE ^{9pT380/Q382K(Ac)}	GLLP TPKK(Ac)SG	1.0 ± 0.1
CycE ^{9pT380/S384K}	GLLP TPPQKSG	4.3 ± 1.3
CycE ^{9pT380/G385K}	GLLP TPPQSK	2.4 ± 0.8
Other phosphopeptides		
CycE ^{9pT380}	GLLP TPPQSG	1.0 ± 0.05
Gcn4 ^{9pT165}	FLPpTPVLED	0.88 ± 0.1
Far1 ^{20pT63}	PKPLNL SKLP SPPPSLKKTA	ND
Sic1 ^{9pT45}	VPVpTPSTTK	ND
Far1 ^{14pT306}	TGEFFQFpTPQEQLI	~25 ± 6
p27 ^{9pT187}	VEQpTPKKPG	ND

Results are the average of at least three individual sets of fluorescence polarization readings. Errors are s.e.m. of all measurements. Values for which saturation binding could not be achieved are indicated as approximate (–). ND indicates no binding detected by fluorescence polarization up to a Skp1–Cdc4 concentration of 10 μM. pT, phosphorylated threonine.

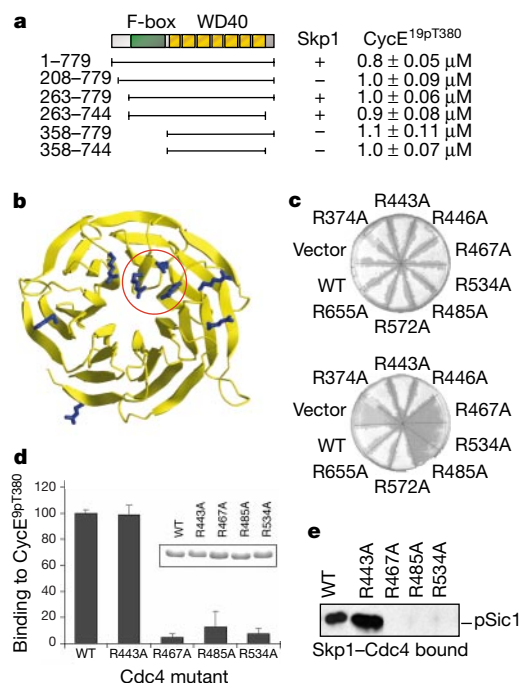


Figure 3 Identification of a CPD binding site on Cdc4. **a**, Equilibrium binding constants for the Cdc4–CycE^{19pT380} interaction determined by fluorescence polarization for a series of Cdc4 deletion mutants. Skp1 binding to Cdc4 was determined by anti-Skp1 immunoblot. **b**, Ribbons diagram of α-carbon chain of the β-propeller structure of the WD40 repeat region of β-transducin. Conserved Arg residues in Cdc4 are superimposed at predicted positions. The cluster of three Arg residues critical for Cdc4 function and peptide binding is circled in red. An alignment of Cdc4 homologues is provided in Supplementary Information. **c**, Mutation of any one of three conserved Arg residues (R467A, R485A, R534A) abolishes Cdc4 function *in vivo*. CDC4 alleles on a TRP1 ARS CEN plasmid were transformed into a cdc4Δ strain containing a CDC4 URA3 CEN plasmid and plated on either Trp⁺Ura[–] (top) or 5-FOA medium (bottom) for 2 days at 30 °C. **d**, **e**, Interaction of Cdc4 WD40 repeat region mutants with CycE^{9pT380} peptide measured by fluorescence polarization (**d**) and phospho-Sic1 detected by anti-Sic1 antibody (**e**). Equal levels of soluble protein were tested (inset).

Thr 380 residue completely eliminated recognition, demonstrating that the heterologous targeting was dependent on phosphorylation. To exclude the possibility that the inserted 19 residues might unduly contort Sic1 and allow recognition in an unnatural context, we substituted the core CPD around either the Thr 45 or Ser 76 sites, without introduction of additional residues. Substitution of the LLpTPP sequence at the Ser 76 site of Sic1^{Op} conferred effective phosphorylation-dependent ubiquitination *in vitro* and degradation *in vivo*, indicating that this single, minimal motif is sufficient for Cdc4 recognition in the context of full-length Sic1 (Fig. 4). The analogous substitution at the Thr 45 site conferred more modest recognition and incomplete ubiquitination and permanent growth arrest on overexpression. Furthermore, conversion of the residues flanking Thr 45 into an optimal CDK recognition sequence did not allow Sic1 recognition *in vitro*, or degradation *in vivo* (Fig. 4). We infer that additional local context effects must influence recognition of the CPD motif. Nonetheless, the fact that an appropriately positioned optimal CPD is sufficient to target Sic1 for degradation *in vivo* raises the question as to why the cell instead relies on multiple, suboptimal CPD sites to eliminate Sic1.

Suboptimal CPDs set a threshold for Sic1 degradation

We sought to test whether the multiple, weak CPD sites in Sic1 are important for biological function. To this end, we introduced a version of Sic1 that lacked seven of the endogenous CDK phosphorylation sites but that incorporated a single, high-affinity LLTPP

motif in place of Ser 76 (Sic1^{2p-S76LLTPP}, hereafter referred to as Sic1^{CPD}), at the chromosomal *SIC1* locus under the control of the endogenous *SIC1* promoter. As Sic1 is needed to link DNA replication to other events at Start⁷, we compared the onset of DNA replication in wild-type and *SIC1*^{CPD} strains on synchronous release from a G1-phase α -factor arrest. To expand the pre-replicative window, cells were grown under suboptimal nutrient conditions. Unlike wild-type cells, which delay all events at Start under these conditions, *SIC1*^{CPD} cells are unable to restrain DNA replication (Fig. 5a). Importantly, Sic1^{CPD} exhibits identical CDK inhibitory activity to that of the wild-type protein (Fig. 5b), indicating that the biological effects of the optimal CPD arise from defective regulation of Sic1 stability. As the Sic1^{CPD} protein proved difficult to detect in wild-type cells, we arrested *SIC1* *cdc28-13* and *SIC1*^{CPD} *cdc28-13* strains at the restrictive temperature of 37°C to inactivate Cdc28 kinases. Sic1^{CPD} was readily detected in the *cdc28-13* arrested cells, and on release to the permissive temperature of 25°C it decayed with a faster onset than wild-type Sic1 (Fig. 5c). Degradation of Sic1^{CPD} was also advanced in early G1 phase populations obtained by elutriation (data not shown). We considered the possibility that targeting of Sic1^{CPD} is rewired to depend on other kinases, such as the CDK enzyme Pho85 (ref. 33), which prefers hydrophobic residues at the +2 position³⁴. However, overexpression of the Pho85-specific cyclin Pcl2 did not affect accumulation of Sic1^{CPD} in *cdc28-13* arrest. Furthermore, Sic1^{CPD} was phosphorylated by Cln2–Cdc28 or Pcl2–Pho85 kinases, and dephosphorylated by the Cdc14 phosphatase, with similar efficiency to that of wild-type Sic1 *in vitro* (data not shown).

As experiments based on cell populations can obscure events at the individual cell level through averaging effects³⁵, we compared the properties of Sic1 and Sic1^{CPD} in single cells by use of green fluorescent protein (GFP) fusions³⁶. Early-G1-phase cells bearing *SIC1*–GFP or *SIC1*^{CPD}–GFP alleles integrated at the chromosomal locus were obtained by centrifugal elutriation. The wild-type Sic1–GFP fusion was detected in all but the largest G1 phase cells, and was absent in all budded cells (Fig. 5d). This result at the level of the single cell is consistent with previous studies on cell cycle regulation of Sic1 abundance in synchronized cultures^{6,7,37}. In contrast, the Sic1^{CPD}–GFP fusion was detected at variable levels and only in a fraction of the elutriated G1 population. As expected, the Sic1^{CPD} mutant was absent from all budded cells. The variable abundance of Sic1^{CPD}–GFP was also evident in G1 cells from unperturbed asynchronous cultures (Fig. 5d). These observations suggest that the optimal CPD site causes precocious and uncoordinated elimination of Sic1.

Premature DNA replication leads to genome instability in both yeast and mammalian cells^{10,14}, presumably because incomplete origin assembly leads to defective genome replication in S phase. Genome stability was measured in wild-type and *SIC1*^{CPD} strains by determining rates of chromosome loss in a colony colour-sectoring assay (Fig. 5e). The rate of chromosome loss was increased by approximately 100-fold in the *SIC1*^{CPD} strain, an effect comparable to that observed for other mutants defective in chromosome transmission³⁸. The pivotal function of Sic1 in the maintenance of genomic integrity is underscored by the massive rate of chromosome loss in a *sic1Δ* strain (Fig. 5e). Sic1 also has a crucial function at the end of mitosis, where it facilitates elimination of Clb–Cdc28 activity to re-establish G1 phase, as shown by the inviability of strains that lack both Sic1 and the APC/C activator Cdh1 (ref. 3). As predicted, we failed to recover *SIC1*^{CPD} *cdh1Δ* double mutants from a cross between *SIC1*^{CPD} and *cdh1Δ* strains (Fig. 5f). This result also indicates that mitotic forms of CDK activity are competent to target Sic1 for degradation through the optimal CPD motif.

Optimal versus suboptimal CPD motifs

The consensus binding site for the WD40 repeats of Cdc4 contains three main determinants. The first is an absolute requirement for

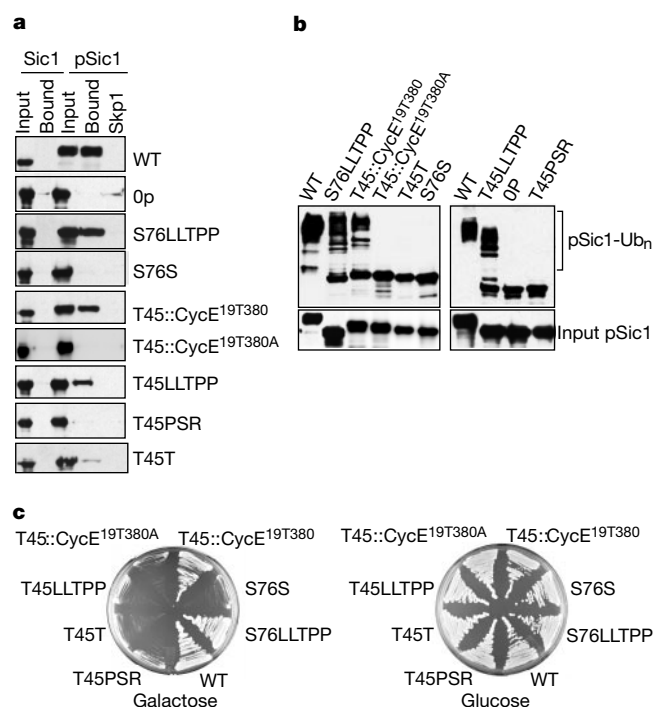


Figure 4 A single, optimal CPD motif is sufficient to target Sic1 to Cdc4. **a**, Capture by Cdc4 of Sic1^{Op} bearing a CycE^{19pT380} insert at Thr 45 or the core CPD motif (LLpTPP) substituted at either Thr 45 or Ser 76. T380A indicates the corresponding non-phosphorylatable insertion; T45PSR indicates a mutant in which the Thr 45 site is converted to an optimal CDK phosphorylation site; S76S and T45T indicate single wild-type sites reintroduced into Sic1^{Op}. GST–Sic1 fusion proteins were either unmodified or phosphorylated by Cln2–Cdc28 and captured on Flag-tagged Skp1–Cdc4 or, as a control, Flag-tagged Skp1 resin. Inputs shown are 40% of total and detection was with anti-Sic1 antibody. **b**, SCF^{Cdc4}-mediated ubiquitination of the same set of Sic1 mutants as in **a**. **c**, Insertion of the CycE^{19pT380} sequence at Thr 45 or substitution of the CPD core motif at Ser 76 restores degradation *in vivo*. *GAL1*–*SIC1* strains were incubated for 2 days at 30°C.

phospho-Ser/Thr followed by a Pro residue. The second is a strong preference for Leu and Ile residues in the -2 position and Pro, Leu or Ile residues in the -1 position. The final requirement is a bias against basic residues in the +2 to +5 positions. Given the minimal experimentally determined CPD, LLpTPP, it is apparent why

inspection of many known phosphorylation sites implicated in targeting various substrates to Cdc4 has failed to yield an obvious consensus sequence. In particular, the nine CDK sites in Sic1 are all non-optimal CPD motifs in that a basic residue is present in the +2 to +5 positions, or a Thr phosphorylation site is replaced with a lower-affinity Ser site, or the -1 and -2 positions lack the preferred hydrophobic residues. Similarly, the eight CDK phosphorylation sites that influence Cdc6 recognition by Cdc4 lack one or more features of an ideal CPD^{20,39,40}. The apparent low affinity of each individual site in Sic1 for Cdc4 explains the requirement for multisite phosphorylation. Given the evidence suggesting a single coincident CPD and substrate-binding site on Cdc4, the strong interaction between phospho-Sic1 and Cdc4 is most easily explained by a high, local concentration of low-affinity binding sites, which effectively drive the equilibrium towards complex formation. The apparent ability of SCF complexes to multimerize might also potentiate high-affinity interactions *in vivo*⁴¹.

There can be no absolute mechanistic requirement for multiple phosphorylation sites in substrate recognition by Cdc4, as it is capable of efficiently capturing substrates that bear a single, high-affinity CPD motif. Despite this observation, it seems probable that most physiological substrates of Cdc4 are recognized by multiple, low-affinity CPD sites. Although Gcn4 degradation depends in part on the strong CPD site centred on Thr 165 (ref. 28), it has recently been shown that multiple CDK sites contribute to Gcn4 degradation⁴²; notably, all of these other sites occur within sub-optimal matches to the CPD. Two other Cdc4 substrates, Far1 and Ash1, also seem to require phosphorylation on multiple sites for efficient Cdc4 recognition (M.T., X.T. and Q. Liu, unpublished data). Similarly, the LRR-containing F-box protein Grr1 targets Cln2 for degradation in a manner that is dependent on multisite phosphorylation^{5,21,43}. The possible generalization of this theme awaits careful investigation of binding interactions between metazoan F-box proteins and their substrates.

Multisite phosphorylation and biological thresholds

Identification of the CPD sequence has uncovered an unexpected mechanism in phosphorylation-dependent protein interactions, in which kinase specificity and phosphopeptide recognition can be antagonistic, as opposed to cooperative, parameters. A dynamic balance between phosphorylation and recognition by the ubiquitination machinery could provide flexibility in substrate degradation to allow fine-tuning of irreversible regulatory switches, such as those that occur in cell cycle transitions. In late G1 phase, a threshold level of Cln1/2–Cdc28 activity is required to activate events associated with Start, including elimination of Sic1 (refs 7, 9). If, on average, phosphorylation on a single site led to the immediate degradation of Sic1, then small fluctuations in CDK activity would erode the ability of the pool of Sic1 to restrain Clb–Cdc28 activity, potentially leading to an uncoordinated onset of DNA replication. Although a single, optimal CPD motif in Sic1 results in recognition by Cdc4, and consequent ubiquitination and degradation, it does not allow precise control of Start, but rather causes precocious S-phase onset and chromosomal instability. That is, the single CPD motif fails to set an appropriate threshold for S phase because it is recognized too efficiently. In contrast, degradation of wild-type Sic1 demands phosphorylation of at least six out of nine CDK sites, thereby imposing a much higher kinase threshold. In this sense, Sic1 integrates Cln–Cdc28 activity, and potentially other G1 phase kinases^{33,44}, to act as a timing device for G1 phase.

Multisite phosphorylation is a common feature of many protein kinase substrates, and yet its biological role has remained enigmatic in most instances. Phosphorylation on multiple residues may help enable events such as multisite docking interactions, integration of different kinase pathways, substrate dephosphorylation, subcellular localization, and protein activity⁴⁵. Perhaps most importantly though, multisite phosphorylation occurs within a cellular

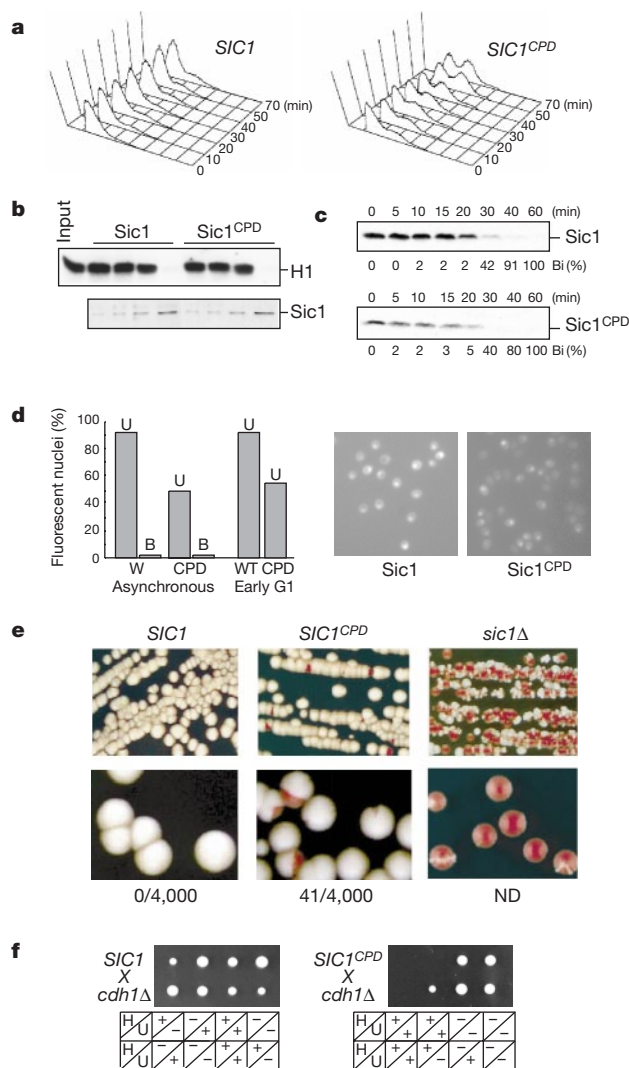


Figure 5 Premature DNA replication and genome instability caused by a single, optimal CPD motif. **a**, Strains bearing integrated wild-type *SIC1* or *SIC1^{2p-ST6LLTPP}* (*SIC1^{CPD}*) alleles were synchronized in G1 phase with α -factor, released into raffinose medium at 25 °C and assessed for total DNA content by FACS analysis. **b**, CDK inhibitory activity of Sic1 and Sic1^{CPD} measured against recombinant Clb5–Cdc28 complexes. At the highest concentration shown, both wild-type Sic1 and Sic1^{CPD} are in slight stoichiometric excess of Clb5. **c**, Premature onset of Sic1^{CPD} protein instability *in vivo*. Strains bearing a *cdc28-13* allele and either wild-type *SIC1* or the *SIC1^{CPD}* allele were incubated at 37 °C for 2 h, released at 25 °C and immunoblotted for total Sic1 protein. Bi, per cent of budded cells. **d**, Analysis of Sic1– and Sic1^{CPD}–GFP fusion protein levels in individual cells. Asynchronous cultures or early-G1-phase populations obtained by centrifugal elutriation were assessed for nuclear fluorescence in unbudded (U) and budded (B) cells. Representative fields are shown for G1-phase populations. **e**, Genome instability caused by the *SIC1^{CPD}* allele. Strains carried a marker chromosome that confers an Ade⁺ phenotype (white colonies); red sectors indicate a chromosome loss event. Primary chromosome loss events were determined by scoring 4,000 individual colonies for red sectors that formed approximately 50% of the colony mass. ND, not determined. **f**, Synthetic lethal interaction between *cdh1Δ* and the *SIC1^{CPD}* allele. H and U indicate deduced His and Ura prototrophy. Of 66 tetrads from the *SIC1^{CPD}–URA3* cross, 46 His⁺ Ura⁺ spore clones did not form colonies, whereas 19 formed small colonies that could not be propagated. Of 31 tetrads from the *SIC1–URA3* cross, two His⁺ Ura⁺ spore clones did not form colonies, whereas 20 formed colonies that could be stably propagated.

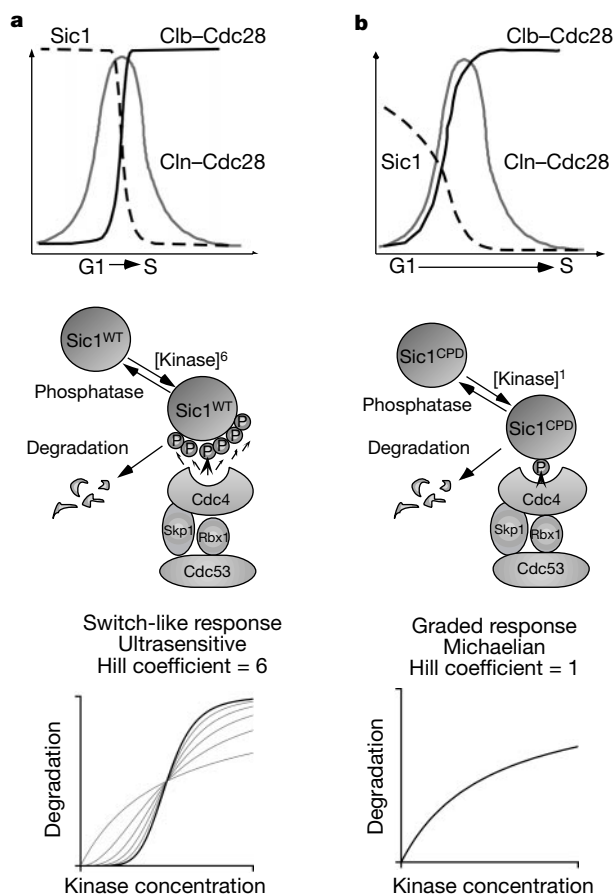


Figure 6 Ultrasensitivity at the G1/S transition. **a**, Multisite phosphorylation of Sic1 sets a threshold for Cln-Cdc28 activity and converts the increase in Cln-Cdc28 into a switch-like response for degradation of Sic1 and onset of Clb-Cdc28 activity. The requirement for six distributive phosphorylation events in Sic1 targeting creates an ultrasensitive response, as modelled by degradation = $([kinase]^{nH})/(K_m + [kinase]^{nH})$, with a Hill coefficient (nH) of six. K_m , Michaelis constant; [kinase], kinase concentration. The theoretical Hill plot shows the predicted response for 1, 2, 3, 4, 5 and 6 phosphorylation sites. **b**, A single, optimal CPD site suffices for efficient targeting of Sic1 to Cdc4, but only with Michaelian kinetics (Hill coefficient = 1) and a hyperbolic response curve, which results in premature onset of degradation and uncoordinated activation of Clb5/6-Cdc28.

environment in which kinases and phosphatases act in dynamic equilibrium, a situation that lends itself to the generation of ultrasensitive biological responses³⁵. The requirement for phosphorylation of Sic1 on at least six sites by definition introduces a highly cooperative step in the degradation pathway (Fig. 6). That is, the forward reaction rate to form the sextuply phosphorylated Sic1 varies as the sixth order of CDK kinase concentration, provided that phosphorylation occurs in a multi-hit distributive manner, as seems to be the case⁴⁶. Additional factors may well enforce the switch-like elimination of Sic1, including inhibition of the phosphatase back-reaction upon Cdc4 binding, operation of CDK kinases at near saturation, and feed-forward action of liberated Clb5/6-Cdc28 on Sic1 (ref. 35). Quantitative measurements of Sic1 degradation rate as a function of relevant kinase, phosphatase, SCF and proteasomal activities should permit mathematical modelling of this crucial step in cell cycle progression, and allow further elucidation of the basis for switch-like biological responses. □

Methods

Yeast strain construction, culture growth, FACS analysis, fluorescence microscopy and plasmid mutagenesis were carried out by standard methods^{38,47}. Strains, plasmids and oligonucleotides used are listed in Tables 1, 2 and 3 of the Supplementary Information. We

carried out recombinant protein production, peptide synthesis, binding assays, kinase assays and ubiquitination reactions essentially as described^{5,48}. Determination of the equilibrium-binding constant was by fluorescence polarization on a Beacon 2000 Fluorescence Polarization System (Pan Vera) at 22 °C using 5 nM fluorescein-labelled probes. Peptide arrays were constructed according to the Spots-synthesis method²⁶ and probed with 1 μM purified Cdc4-Skp1 complex. Complete details of all methods are provided in the Supplementary Information.

Received 22 June; accepted 12 August 2001.

- Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479 (1998).
- Varshavsky, A. Naming a targeting signal. *Cell* **64**, 13–15 (1991).
- Tyres, M. & Jorgensen, P. Proteolysis and the cell cycle: with this RING I do thee destroy. *Curr. Opin. Genet. Dev.* **10**, 54–64 (2000).
- Bai, C. *et al.* SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* **86**, 263–274 (1996).
- Skowyra, D., Craig, K. L., Tyers, M., Elledge, S. J. & Harper, J. W. F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **91**, 209–219 (1997).
- Schwob, E., Bohm, T., Mendenhall, M. D. & Nasmyth, K. The B-type cyclin kinase inhibitor p40^{SIC1} controls the G1 to S transition in *S. cerevisiae*. *Cell* **79**, 233–244 (1994).
- Schneider, B. L., Yang, Q. H. & Futcher, A. B. Linkage of replication to start by the Cdk inhibitor Sic1. *Science* **272**, 560–562 (1996).
- Feldman, R. M., Correll, C. C., Kaplan, K. B. & Deshaies, R. J. A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* **91**, 221–230 (1997).
- Verma, R. *et al.* Phosphorylation of Sic1p by G1 Cdk required for its degradation and entry into S phase. *Science* **278**, 455–460 (1997).
- Nugroho, T. T. & Mendenhall, M. D. An inhibitor of yeast cyclin-dependent protein kinase plays an important role in ensuring the genomic integrity of daughter cells. *Mol. Cell. Biol.* **14**, 3320–3328 (1994).
- Amati, B. & Vlach, J. Kip1 meets SKP2: new links in cell-cycle control. *Nature Cell Biol.* **1**, E91–E93 (1999).
- Winston, J. T., Chu, C. & Harper, J. W. Culprits in the degradation of cyclin E apprehended. *Genes Dev.* **13**, 2751–2757 (1999).
- Sheaff, R. J., Groudine, M., Gordon, M., Roberts, J. M. & Clurman, B. E. Cyclin E-CDK2 is a regulator of p27^{Kip1}. *Genes Dev.* **11**, 1464–1478 (1997).
- Spruck, C. H., Won, K. A. & Reed, S. I. Deregulated cyclin E induces chromosome instability. *Nature* **401**, 297–300 (1999).
- Maniatis, T. A ubiquitin ligase complex essential for the NF-κB, Wnt/Wingless, and hedgehog signaling pathways. *Genes Dev.* **13**, 505–510 (1999).
- Yaron, A. *et al.* Identification of the receptor component of the IκBα-ubiquitin ligase. *Nature* **396**, 590–594 (1998).
- Carrano, A. C., Eytan, E., Hershko, A. & Pagano, M. SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27. *Nature Cell Biol.* **1**, 193–199 (1999).
- Tsvetkov, L. M., Yeh, K. H., Lee, S. J., Sun, H. & Zhang, H. p27^{Kip1} ubiquitination and degradation is regulated by the SCF^{Skp2} complex through phosphorylated Thr187 in p27. *Curr. Biol.* **9**, 661–664 (1999).
- Pawson, T. & Nash, P. Protein-protein interactions define specificity in signal transduction. *Genes Dev.* **14**, 1027–1047 (2000).
- Drury, L. S., Perkins, G. & Diffley, J. F. The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast. *EMBO J.* **16**, 5966–5976 (1997).
- Lanker, S., Valdivieso, M. H. & Wittenberg, C. Rapid degradation of the G1 cyclin Cln2 induced by CDK-dependent phosphorylation. *Science* **271**, 1597–1601 (1996).
- Schneider, B. L. *et al.* Yeast G1 cyclins are unstable in G1 phase. *Nature* **395**, 86–89 (1998).
- Henchoz, S. *et al.* Phosphorylation- and ubiquitin-dependent degradation of the cyclin-dependent kinase inhibitor Far1p in budding yeast. *Genes Dev.* **11**, 3046–3060 (1997).
- Clurman, B. E., Sheaff, R. J., Thress, K., Groudine, M. & Roberts, J. M. Turnover of cyclin E by the ubiquitin-proteasome pathway is regulated by Cdk2 binding and cyclin phosphorylation. *Genes Dev.* **10**, 1979–1990 (1996).
- Won, K. A. & Reed, S. I. Activation of cyclin E/CDK2 is coupled to site-specific autophosphorylation and ubiquitin-dependent degradation of cyclin E. *EMBO J.* **15**, 4182–4193 (1996).
- Frank, R. Spot-synthesis: an easy technique for positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron* **48**, 9217–9232 (1992).
- Songyang, Z. *et al.* Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Curr. Biol.* **4**, 973–982 (1994).
- Meimoun, A. *et al.* Degradation of the transcription factor Gcn4 requires the kinase Pho85 and the SCF^{CDC4} ubiquitin-ligase complex. *Mol. Biol. Cell* **11**, 915–927 (2000).
- McKinney, J. D. & Cross, F. R. FAR1 and the G1 phase specificity of cell cycle arrest by mating factor in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 2509–2516 (1995).
- Gartner, A. *et al.* Pheromone-dependent G1 cell cycle arrest requires Far1 phosphorylation, but may not involve inhibition of Cdc28-Cln2 kinase, *in vivo*. *Mol. Cell. Biol.* **18**, 3681–3691 (1998).
- Yaffe, M. B. & Elia, A. E. Phosphoserine/threonine-binding domains. *Curr. Opin. Cell Biol.* **13**, 131–138 (2001).
- Wall, M. A. *et al.* The structure of the G protein heterotrimer Gα1β1γ2. *Cell* **83**, 1047–1058 (1995).
- Nishizawa, M., Kawasumi, M., Fujino, M. & Toh-e, A. Phosphorylation of Sic1, a cyclin-dependent kinase (Cdk) inhibitor, by Cdk including Pho85 kinase is required for its prompt degradation. *Mol. Biol. Cell* **9**, 2393–2405 (1998).
- O'Neill, E. M., Kaffman, A., Jolly, E. R. & O'Shea, E. K. Regulation of PHO4 nuclear localization by the PHO80-PHO85 cyclin-CDK complex. *Science* **271**, 209–212 (1996).
- Ferrell, J. E. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends Biochem. Sci.* **21**, 460–466 (1996).
- Cormack, B. Green fluorescent protein as a reporter of transcription and protein localization in fungi. *Curr. Opin. Microbiol.* **1**, 406–410 (1998).
- Donovan, J. D., Toyn, J. H., Johnson, A. L. & Johnston, L. H. P40^{SDR25}, a putative CDK inhibitor, has a role in the M/G1 transition in *Saccharomyces cerevisiae*. *Genes Dev.* **8**, 1640–1653 (1994).

38. Spencer, F., Gerring, S. L., Connelly, C. & Hieter, P. Mitotic chromosome transmission fidelity mutants in *Saccharomyces cerevisiae*. *Genetics* **124**, 237–249 (1990).
39. Elsasser, S., Chi, Y., Yang, P. & Campbell, J. L. Phosphorylation controls timing of Cdc6p destruction: a biochemical analysis. *Mol. Biol. Cell* **10**, 3263–3277 (1999).
40. Drury, L. S., Perkins, G. & Diffley, J. F. The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. *Curr. Biol.* **10**, 231–240 (2000).
41. Wolf, D. A., McKeon, F. & Jackson, P. K. F-box/WD-repeat proteins Pop1p and Sud1p/Pop2p form complexes that bind and direct the proteolysis of Cdc18p. *Curr. Biol.* **9**, 373–376 (1999).
42. Chi, Y. *et al.* Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase. *Genes Dev.* **15**, 1078–1092 (2001).
43. Patton, E. E., Willems, A. R. & Tyers, M. Combinatorial control in ubiquitin-dependent proteolysis: don't Skp the F-box hypothesis. *Trends Genet.* **14**, 236–243 (1998).
44. Dirick, L., Goetsch, L., Ammerer, G. & Byers, B. Regulation of meiotic S phase by Ime2 and a Clb5,6-associated kinase in *Saccharomyces cerevisiae*. *Science* **281**, 1854–1857 (1998).
45. Cohen, P. The regulation of protein function by multisite phosphorylation: a 25 year update. *Trends Biochem. Sci.* **25**, 596–601 (2000).
46. Annan, R. S., Huddleston, M. J., Verma, R., Deshaies, R. J. & Carr, S. A. A multidimensional electrospray MS-based approach to phosphopeptide mapping. *Anal. Chem.* **73**, 393–404 (2001).
47. Guthrie, C. & Fink, G. R. (eds) *Methods in Enzymology* (Academic, San Diego, 1991).
48. van der Geer, P., Wiley, S., Gish, G. D. & Pawson, T. The Shc adaptor protein is highly phosphorylated at conserved, twin tyrosine residues (Y239/240) that mediate protein-protein interactions. *Curr. Biol.* **6**, 1435–1444 (1996).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank G. Gish for helpful discussions and initial peptide synthesis; A. Breitzkreutz for assistance with phosphorylation reactions; A. Willems for assistance with modelling; and D. Durocher, A. Amon and L. Harrington for critical comments on the manuscript. This work was supported by grants from the Canadian Institutes of Health Research (CIHR), the National Cancer Institute of Canada (NCIC), the Protein Engineering Network of Centers of Excellence (PENCE) and MDS/Sciex (F.S., T.P. and M.T.), the Human Frontiers Science Program (M.T.), and from the National Institutes of Health (F.G. and M.D.M.). P.N. and X.T. are recipients of CIHR postdoctoral fellowship awards, T.P. is a Distinguished Scientist of the CIHR, M.T. is a Canada Research Chair in Biochemistry and F.S. is an NCIC Scientist.

Correspondence and requests for materials should be addressed to T.P. (e-mail: pawson@mshri.on.ca) or M.T. (e-mail: tyers@mshri.on.ca). The protein interactions are available at the Biomolecular Interaction Network Database (accession numbers 1436, 1437 and 1438).

An unusually massive stellar black hole in the Galaxy

J. Greiner*, J. G. Cuby† & M. J. McCaughrean*

* Astrophysical Institute Potsdam, 14482 Potsdam, Germany

† ESO, Alonso de Córdova 3107, Santiago 19, Chile

The X-ray source known as GRS1915+105 belongs to a group dubbed 'microquasars'^{1,2}. These objects are binary systems which sporadically eject matter at speeds that appear superluminal, as is the case for some quasars. GRS1915+105 is also one of only two known binary sources thought to contain a maximally spinning black hole³. Determining the basic parameters of GRS1915+105, such as the masses of the components, will help us to understand jet formation in this system, as well as providing links to other objects which exhibit jets. Using X-ray data, indirect methods^{4,5} have previously been used to infer a variety of masses for the accreting compact object in the range 10–30 solar masses ($M_{\odot}$). Here we report a direct measurement of the orbital period and mass function of GRS1915+105, which allow us to deduce a mass of $14 \pm 4 M_{\odot}$ for the black hole. Black holes with masses $>5\text{--}7 M_{\odot}$ challenge the conventional picture of black-hole formation in binary systems^{6–9}. Based on the mass estimate, we interpret the distinct X-ray variability of GRS1915+105 as arising from instabilities in an accretion disk that is dominated by radiation pressure, and radiating near the Eddington limit (the point where radiation pressure supports matter against gravity). Also, the mass estimate constrains most models which relate observable X-ray properties to the spin of black holes in microquasars.

GRS1915+105 is located in the Galactic plane at a distance of $\sim 11\text{--}12$ kpc (refs 10, 11) and suffers a large extinction of 25–30 mag in the visual band. Spectroscopic observations in the near-infrared H and K bands identified absorption features from the atmosphere of the companion (mass-donating star) in the

GRS1915+105 binary¹². The detection of ^{12}CO and ^{13}CO band heads plus a few metallic absorption lines suggested a K-M spectral type and luminosity class III (giant).

The presence of these band-head features led us to carry out follow-up medium-resolution spectroscopy in the 2.29–2.41 μm wavelength range using the Very Large Telescope (VLT)-Antu equipped with Infrared Spectrometer and Array Camera ISAAC, between April and August 2000 (Fig. 1). Radial velocities were measured for the 16 individual spectra by cross-correlation of the major CO band heads, and a period analysis was carried out (Fig. 2). The periodogram shows a clear peak at a period of 33.5 days (top panel) which we interpret as the orbital period P_{orb} of the binary system. The velocity amplitude is measured to be $K_d = 140 \pm 15 \text{ km s}^{-1}$ (lower panel). Figure 1 shows that the infrared flux is dominated by light from the accretion flow or jet, rather than from the secondary star. There is thus a possibility that phase-dependent changes in the continuum near the absorption features may result in an additional source of systematic error in the measured value of K_d . The measured parameters allow us to determine the mass function $f(M)$, that is, the observational lower limit to the mass of the compact object:

$$f(M) \equiv \frac{(M_c \sin i)^3}{(M_c + M_d)^2} = \frac{P_{\text{orb}} K_d^3}{2\pi G} = 9.5 \pm 3.0 M_{\odot} \quad (1)$$

In order to determine the true mass of the black hole, M_c , estimates of the donor mass M_d and the orbital inclination i are required. (G is the gravitational constant.) The K-M III classification at a first approximation implies a mass of $M_d = 1.2 \pm 0.2 M_{\odot}$ for the donor¹². Because of the high mass-loss of the donor (needed to explain the large X-ray luminosity), the donor is almost certainly less luminous than a non-interacting star of the same spectral type. This in turn would imply a larger donor mass, thus making the black-hole mass estimate (see below) a lower limit when using $M_d = 1.2 M_{\odot}$. The orbital inclination of the GRS1915+105 binary can be deduced from the orientation of the jet, which in turn is derived from the brightness and the velocities of both the approaching and receding blobs^{10,11}. This angle of $70^\circ \pm 2^\circ$ was observed to be constant over several years, indicating no measurable precession,

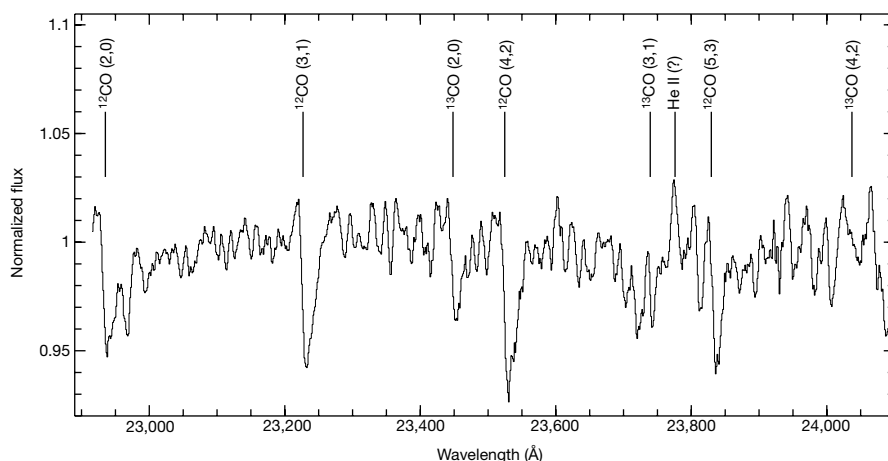


Figure 1 Mean K-band spectrum of GRS1915+105. It was obtained at the ESO VLT-Antu telescope, using the short-wavelength (0.9–2.5 μm) arm of ISAAC, equipped with a $1,024 \times 1,024$ pixel Rockwell HgCdTe array with an image scale of 0.147'' per pixel. The medium resolution grating (1.2 Å per pixel in the K band) was used, which yielded a spectral resolution of $\sim 3,000$ with a 1'' slit. Exposures of GRS1915+105 consisted of eight 250-s individual exposures which were dithered along the slit by $\pm 10''$. In order to correct for atmospheric absorption, the nearby star HD179913 (A0 V) was observed either before or after each exposure. The initial data reduction steps such as bias subtraction, flatfielding and co-adding were performed within the Eclipse package²⁶. The extraction

and wavelength calibration was done using an optimal extraction routine within the MIDAS package²⁹. The spectrum shown here is the sum of 5 exposures with 167 min total integration time. Band heads of CO are marked with vertical lines and the numbers in brackets denote the energy levels of the transition. The presence of the ^{13}CO isotope and the equivalent width ratio of ^{12}CO to ^{13}CO suggests a classification of the donor as a late-type giant. The small width and faintness of the CO band heads imply that the donor contributes only a few per cent to the total K-band brightness (see ref. 12 for details) of the binary system GRS1915+105.

and justifies the assumption that the jet is perpendicular to the accretion disk and orbital plane. Knowing the inclination i and a lower limit of the donor mass, we can now solve equation (1) for the mass of the accreting compact object (Fig. 3), finding $M_c = 14 \pm 4 M_\odot$. Table 1 summarizes all orbital parameters of GRS1915+105. Even after accounting for the relatively large error dominated by the determination of the velocity amplitude K_d , GRS1915+105 is the Galactic low-mass X-ray binary with the largest known mass function and the largest known mass of its compact object. Previous record holders were V404 Cyg with $f(M) = 6.07 \pm 0.05$, $M_c = 7\text{--}10 M_\odot$ (ref. 13) and XTE J1118+480 with $f(M) = 6.00 \pm 0.36$, $M_c = 6.5\text{--}10 M_\odot$ (ref. 14).

The mass of the black hole in GRS1915+105 has several implications for our understanding of the physics of microquasars, as well as some broader astrophysical concepts. Most importantly, the formation of a $14 M_\odot$ black hole in a low-mass binary poses a challenge for binary evolution models. Stellar evolution of stars in a binary system proceeds differently from single star evolution primarily due to the mass transfer between the system components and/or common-envelope phases. There are, in general, two different paths for the black-hole formation in a binary system. First, the progenitor system could be wide, and during the common-envelope phase the low-mass (main sequence) star of $\sim M_\odot$ will spiral into the envelope of the massive giant (progenitor of the black hole), causing the orbit to shrink^{9,15}. Based on our measured system parameters (Table 1), the deduced orbital separation of the binary components in GRS1915+105 is $108 \pm 4 R_\odot$ ($R_\odot$ is the radius of the Sun, 6.9×10^5 km). Thus, orbital contraction through a common-envelope phase caused by the expansion of the massive progenitor to typically $\geq 1,000 R_\odot$, is conceivable for GRS1915+105. Second, the evolution could start with a progenitor system smaller than today, provided the binary component interaction is delayed until after helium burning has ceased⁶. In this case, the time between the

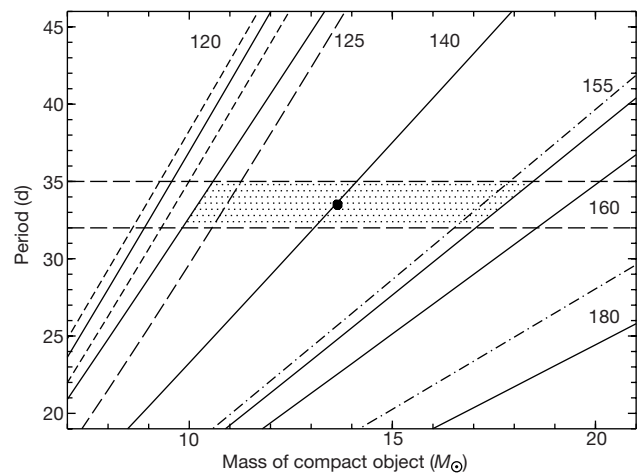


Figure 3 Black-hole mass constraints for GRS1915+105. The relation of orbital period versus mass of the black hole is plotted for various velocity amplitudes K_d (solid lines) in km s^{-1} . We assumed an orbital inclination of 70° and a mass of the donor of $1.2 M_\odot$. The horizontal long-dashed lines indicate the boundaries of the period uncertainty, and the radial velocity range is $125\text{--}155 \text{ km s}^{-1}$. Thus, the dotted region shows the allowed parameter space, leading to a mass of the accreting compact object of $14 \pm 4 M_\odot$. The implied Roche lobe size of the donor star is $21 \pm 4 R_\odot$, in good agreement with the size of a K-M giant which is thus likely to fill its Roche lobe. The uncertainty in the mass of the donor is shown for $K_d = 120 \text{ km s}^{-1}$ where the slanted dashed lines correspond to 1.0, 1.4 and $2.5 M_\odot$, respectively (from left to right). While the formal uncertainty in the orbital inclination is only 2° , we show the effect of relaxing the assumption that the jet is perpendicular to the orbital plane by showing for the $K_d = 160 \text{ km s}^{-1}$ case the corresponding curves using $i = 79^\circ$ (at which angle eclipses would set in; left dash-dot curve) and $i = 61^\circ$ (right dash-dot curve). When relaxing the assumptions and using the extremes, the mass range would be $8\text{--}24 M_\odot$.

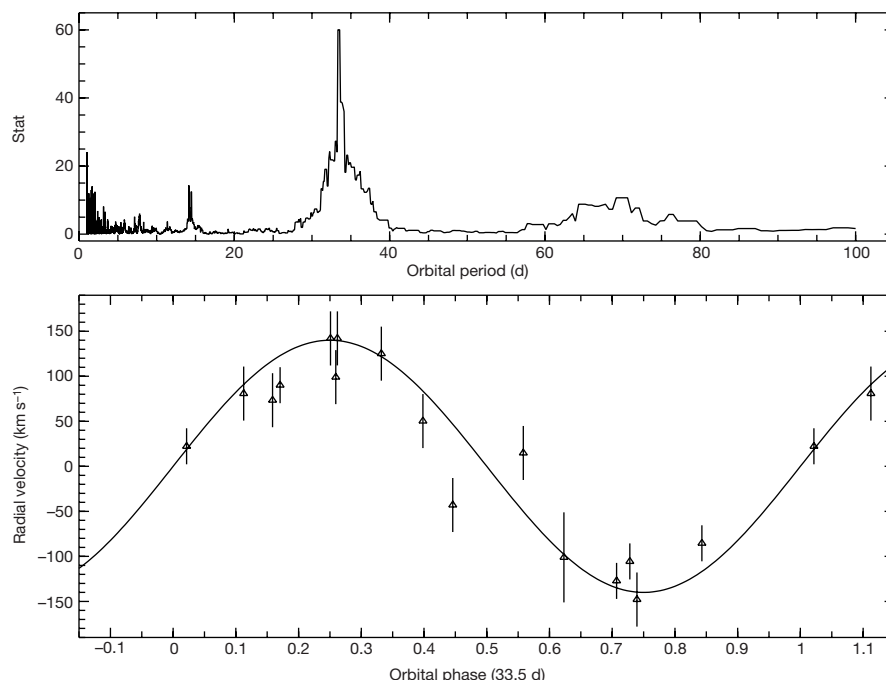


Figure 2 Period analysis of the velocity variation of the four CO band heads. Radial velocities were measured for the individual spectra by cross-correlation of the major CO band heads, using as template a spectrum of the K2III star HD202135 taken with the same setting. Top: Scargle periodogram after heliocentric correction of the individual measurements. Bottom: radial velocity curve folded over the best-fit period of $P_{\text{orb}} = 33.5$ days. The semi-amplitude of the velocity curve K_d is $140 \pm 15 \text{ km s}^{-1}$.

Distortions of the radial velocity curve due to X-ray heating (see, for example, ref. 27) are expected to be unimportant because of the long orbital period. The systemic velocity is $\gamma = -3 \pm 10 \text{ km s}^{-1}$ which implies that based on the Galactic rotation curve²⁸ the kinematic distance (d) of GRS1915+105 is $d = 12.1 \pm 0.8 \text{ kpc}$, intermediate between earlier estimates^{10,11}. 'Stat' is a measure of the significance of the period.

wind phase and the core-collapse is short, and black-hole masses in the 5–10 $M_{\odot}$ range are plausible when the initial helium-star progenitor is in the 10–25 $M_{\odot}$ range, corresponding to initial primaries with 25–45 $M_{\odot}$ (refs 8, 9). The total mass that is finally lost depends on the evolution of the two progenitor star radii, and it remains to be shown whether black-hole masses above 10 $M_{\odot}$ can be achieved. In order to produce even higher black-hole masses, the progenitor might have been a massive Wolf–Rayet star. However, Wolf–Rayet stars have a much larger wind-loss rate, and it is therefore unclear whether higher progenitor masses indeed will lead to higher final black-hole masses. An alternative way of producing high mass ($\geq 10 M_{\odot}$) black holes may be to invoke hierarchical triples as progenitors¹⁶.

Knowledge of the mass of the accretor in GRS1915+105 also yields insight into the rapid and large-amplitude X-ray variability seen in this source¹⁷, which occurs near or even above the Eddington accretion rate limit $\dot{M}_{\text{Edd}}$. Such high accretion rates are not reached by other canonical black-hole transients (for example, GRO J1655–40) which usually operate in the 0.1–0.2 $\dot{M}_{\text{Edd}}$ range, at which their accretion disks are probably gas pressure dominated, and thus viscously and thermally stable. The high $\dot{M}/\dot{M}_{\text{Edd}}$ ratio in GRS1915+105 suggests that its inner accretion disk is radiation pressure dominated, which in turn makes the inner disk quasi-spherical and thermally unstable. This property provides a clue to the X-ray variability in GRS1915+105 (ref. 17). It is tempting to conclude that jet ejection occurs because the black hole can not accept this copious supply of matter. But jet ejection also occurs in other sources (for example, at 0.2 $\dot{M}_{\text{Edd}}$ in GRO J1655–40), and thus near/super-Eddington accretion cannot be the determining factor for relativistic jets.

Finally, if the black-hole mass in GRS1915+105 is indeed no larger than 18 $M_{\odot}$ (Fig. 3), we can place constraints on the black-hole spin in GRS1915+105 and GRO J1655–40. Previously, information on the black-hole spin has been deduced from two completely different sources. First, accretion disks around a (prograde) spinning black hole extend farther down towards the black hole, and thus allow the temperature of the disk to be higher. Both GRS1915+105 and GRO J1655–40 exhibit a thermal component in their X-ray spectra which is unprecedentedly high when compared to all other black-hole transients (during outbursts). It has thus been argued that this is due to their black-hole spin (most black-hole transients have non-rotating black holes³). Second, several black-hole binaries, including GRS1915+105 and GRO J1655–40, show near-stable quasi-periodic oscillations (QPOs) in their X-ray emission. The frequency of oscillation, f , is 300 Hz in GRO J1655–40 (ref. 18) and 67 Hz in GRS1915+105 (ref. 5). Most of the models proposed^{19–22} to explain these QPOs depend upon the spin of the accreting black hole. The black-hole mass of GRS1915+105 (M_{c} in units of $M_{\odot}$) makes the deduction of the black-hole spin in ref. 3 inconsistent with any of these models. (1) If the QPO frequency is associated with the keplerian motion at the last stable orbit around a (non-rotating) black hole according to the simple relation $f = 2.2/M_{\text{c}}$, where f is in kHz, we find agreement with the optically determined mass for GRO J1655–40, but an error

of a factor of 2 for GRS1915+105; that is, the QPO frequency does not scale linearly with the mass of the black hole. (2) If the QPO frequency is associated with the trapped g-mode (diskoseismic) oscillations near the inner edge of the accretion disk^{19,20}, the model would require a nearly maximally spinning black hole in GRO J1655–40, and a non-spinning black hole in GRS1915+105. (3) Similarly, if associated with the relativistic dragging of inertial frames around a spinning black hole²¹ which would cause the accretion disk to precess, the implied specific angular momentum a (spin) of the black hole in GRS1915+105 would be $a \approx 0.8$, considerably lower than the $a \approx 0.95$ deduced for GRO J1655–40. The implications of both of these models are in conflict with the nearly identical accretion-disk temperatures for both sources which in turn requires a larger spin for GRS1915+105 (ref. 3). (4) If the QPO frequency is associated with oscillations related to a centrifugal barrier in the inner part of the accretion disk²², the product of QPO frequency and black-hole mass is predicted to be proportional to the accretion rate, implying that the accretion rate in GRO J1655–40 should be a factor of ~ 10 larger than in GRS1915+105. This is certainly not the case.

Thus, none of these four models provides a satisfactory solution if we adopt the interpretation that the high accretion-disk temperatures are a measure of the black-hole spin³. But if this interpretation is dropped and the spin becomes a free parameter, then the first three models could be applicable. The deduction of accurate accretion-disk temperatures from the applied disk model has also been questioned on other grounds^{23,24}. Additionally, there exist alternative, so-called slim disk models, which also reproduce high-temperature disks for non-rotating black holes²⁵. $\square$

Received 14 June; accepted 12 October 2001.

- Mirabel, I. F. & Rodríguez, L. F. Microquasars in our Galaxy. *Nature* **392**, 673–676 (1998).
- Greiner, J. *Microquasars in Cosmic Explosions*, Proc. 10th Annu. Astrophys. Conf. (eds Holt, S. & Zhang, W. W.) 307–316 (AIP 522, American Institute of Physics, Melville, New York, 2000).
- Zhang, N. S., Cui, W. & Chen, W. Black hole spin in X-ray binaries: Observational consequences. *Astrophys. J.* **482**, L155–L158 (1997).
- Greiner, J., Morgan, E. H. & Remillard, R. A. RXTE Spectroscopy of GRS 1915+105. In Galactic Sources with Relativistic Jets (eds Ogley, R. N. & Bell Burnell, J.) *New Astr. Rev.* **42**, 597–600 (1998).
- Morgan, E. H., Remillard, R. A. & Greiner, J. RXTE observations of QPOs in the black hole candidate GRS 1915+105. *Astrophys. J.* **482**, 993–1010 (1997).
- Brown, G. E., Lee, C.-H. & Bethe, H. A. The formation of high-mass black holes in low-mass X-ray binaries. *New Astron.* **4**, 313–323 (1999).
- Wellstein, S. & Langer, N. Implications of massive close binaries for black hole formation and supernovae. *Astron. Astrophys.* **350**, 148–162 (1999).
- Brown, G. E., Lee, C.-H. & Tauris, T. M. Formation and evolution of black hole X-ray transient systems. *New Astron.* **6**(7), 457–470 (2001).
- Kalogera, V. in *Evolution of Binary and Multiple Star Systems* (eds Podsiadlowski, P. et al.) (ASP Conf. Ser., Astronomical Society of the Pacific, San Francisco, in the press); also as preprint astro-ph/0012064 at (<http://xxx.lanl.gov>) (2001).
- Mirabel, I. F. & Rodríguez, L. F. A superluminal source in the galaxy. *Nature* **371**, 46–48 (1994).
- Fender, R. P. et al. MERLIN observations of relativistic ejections from GRS 1915+105. *Mon. Not. R. Astron. Soc.* **304**, 865–876 (1999).
- Greiner, J., Cuby, J. G., McCaughrean, M. J., Castro-Tirado, A. J. & Mennickent, R. E. Identification of the donor in the X-ray binary GRS 1915+105. *Astron. Astrophys.* **373**, L37–L40 (2001).
- Shahbaz, T., Bandyopadhyay, R., Charles, P. A. & Naylor, T. Infrared spectroscopy of V404 Cygni: limits on the accretion disc contamination. *Mon. Not. R. Astron. Soc.* **282**, 977–981 (1996).
- McClintock, J. E. et al. A black hole of $> 6 M_{\odot}$ in the X-ray nova XTE J1118+480. *Astrophys. J.* **551**, L147–L150 (2001).
- Kalogera, V. Donor stars in black hole X-ray binaries. *Astrophys. J.* **521**, 723–734 (1999).
- Eggleton, P. P. & Verbunt, F. Triple star evolution and the formation of short-period, low mass X-ray binaries. *Mon. Not. R. Astron. Soc.* **220**, 13p–18p (1986).
- Greiner, J., Morgan, E. H. & Remillard, R. A. Rossi X-Ray Timing Explorer observations of GRS 1915+105. *Astrophys. J.* **473**, L107–L110 (1996).
- Remillard, R. A., Morgan, E. H., McClintock, J. E., Bailyn, C. D. & Orosz, J. A. RXTE observations of 0.1–300 Hz quasi-periodic oscillations in the microquasar GRO J1655–40. *Astrophys. J.* **522**, 397–412 (1999).
- Okazaki, A. T., Kato, S. & Fukue, J. Global trapped oscillations of relativistic accretion disks. *Publ. Astron. Soc. Jpn* **39**, 457–473 (1987).
- Nowak, M. A., Wagoner, R. V., Begelman, M. C. & Lehr, D. E. The 67 Hz feature in the black hole candidate GRS 1915+105 as a possible “diskoseismic” mode. *Astrophys. J.* **477**, L91–L94 (1997).
- Cui, W., Zhang, S. N. & Chen, W. Evidence for frame-dragging around spinning black holes in X-ray binaries. *Astrophys. J.* **492**, L53–L56 (1998).
- Titarchuk, L., Lapidus, I. & Muslimov, A. Mechanisms for high-frequency quasi-periodic oscillations in neutron star and black hole binaries. *Astrophys. J.* **499**, 315–328 (1998).
- Sobczak, G. J., McClintock, J. E., Remillard, R. E., Bailyn, C. D. & Orosz, J. A. RXTE spectral

Table 1 Spectroscopic orbital parameters of GRS1915+105

Parameter	Result
T_0 (UT)*	2000 May 02 00:00
T_0 (heliocentric)*	HJD 2451666.5 $\pm$ 1.5
γ (km s ^{−1})	−3 $\pm$ 10
K_d (km s ^{−1})	140 $\pm$ 15
P_{orb} (d)	33.5 $\pm$ 1.5
$f(M)$ ($M_{\odot}$)	9.5 $\pm$ 3.0
M_d ($M_{\odot}$)	1.2 $\pm$ 0.2
M_c ($M_{\odot}$)†	14 $\pm$ 4

* Time of blue-to-red crossing.

† Using an inclination angle $i = 70^\circ \pm 2^\circ$ (refs 10,11) (see Fig. 3 legend).

- observations of the 1996–1997 outburst of the microquasar GRO J1655–40. *Astrophys. J.* **520**, 776–787 (1999).
24. Merloni, A., Fabian, A. C. & Ross, R. R. On the interpretation of the multicolour disc model for black hole candidates. *Mon. Not. R. Astron. Soc.* **313**, 193–197 (2000).
25. Watarai, K., Fukue, J., Takeuchi, M. & Mineshige, S. Galactic black-hole candidates shining at the Eddington luminosity. *Publ. Astron. Soc. Jpn* **52**, 133–141 (2000).
26. Devillard, N. Eclipse users guide. (<http://www.eso.org/projects/aot/eclipse/eug/index.html>) (2000).
27. Phillips, S. N., Shahbaz, T. & Podsiadlowski, Ph. The outburst radial velocity curve of X-ray Nova Scorpii 1994 (=GRO J1655–40): a reduced mass for the black hole? *Mon. Not. R. Astron. Soc.* **304**, 839–844 (1999).
28. Fich, M., Blitz, L. & Stark, A. A. The rotation curve of the Milky Way to 2 R₀. *Astrophys. J.* **342**, 272–284 (1989).
29. Warmels, R. H. in *Astronomical Data Analysis Software and Systems I* 115–119 (PASP Conf. Ser. 25, Astronomical Society of the Pacific, San Francisco, 1991).

Acknowledgements

This work is based on observations collected at the European Southern Observatory, Chile.

Correspondence and requests for materials should be addressed to J.G.
(e-mail: jgreiner@aip.de).

Atomic structure holography using thermal neutrons

B. Sur*, R. B. Rogge†, R. P. Hammond‡, V. N. P. Anghel* & J. Katsaras†

* Atomic Energy of Canada Limited, Chalk River Laboratories, Chalk River, Ontario K0J 1J0, Canada

† National Research Council, Steacie Institute for Molecular Sciences, Chalk River, Ontario K0J 1J0, Canada

‡ Department of Physics and Astronomy, McMaster University, Hamilton, Ontario L8S 4M1, Canada

The idea of atomic-resolution holography has its roots in the X-ray work of Bragg¹ and in Gabor's electron interference microscope². Gabor's lensless microscope was not realized in his time, but over the past twelve years there has been a steady increase in the number of reports on atomic-resolution holography. All of this work involves the use of electrons^{3–6} or hard X-rays^{7–11} to produce the hologram. Neutrons are often unique among scattering probes in their interaction with materials: for example, the relative visibility of hydrogen and its isotopes is a great advantage in the study of polymers and biologically relevant materials. Recent work¹² proposed that atomic-resolution holography could be achieved with thermal neutrons. Here we use monochromatic thermal neutrons, adopting the inside-source concept of Szöke¹³, to image planes of oxygen atoms located above and below a single hydrogen atom in the oxide mineral simpsonite¹⁴.

The inside-source concept¹⁵ uses the atoms making up the sample as independent sources of coherent illumination. In the case of X-ray holography and certain types of electron holography (such as auger holography), the radiation is generated by atomic de-excitation that propagates in the form of nearly spherical waves. This radiation can pass through the sample unperturbed (reference beam), or be scattered by the surrounding atoms (object beam). The interference between the reference and object beams produces the hologram. Similarly, incoherent scattering of thermal neutrons produces spherical waves. Hydrogen has a large incoherent neutron-scattering cross-section: $\sigma_i = 80$ barns. Unlike forms of electromagnetic radiation that interact primarily with electrons, neutrons interact directly with nuclei and are, to first order, equally capable of imaging light atoms and heavy atoms. Consequently, atomic-resolution holographic images can be obtained of materials comprised entirely of light atoms (for example, hydrocarbons). Both

holography and conventional direct methods solve the phase problem in crystallography. Inside-source holography has the added advantage that only orientational order is required, that is, translational symmetry is not necessary. Holography is therefore also suitable for imaging non-crystalline materials such as nematic liquid crystals. Because of the large σ_i of hydrogen, thermal neutron holography should be particularly effective for the *ab initio* solution of organic macromolecule structures.

The sample used to demonstrate the feasibility of neutron holography was a single crystal of natural simpsonite, a rare oxide mineral of aluminium and tantalum with chemical formula Al₄Ta₃O₁₃(OH). Simpsonite was first discovered in western Australia¹⁴ in 1939 and was examined using X-ray diffraction¹⁶ the same year. More recently¹⁷ X-ray diffraction has been used to refine the crystal structure in the trigonal space group *P*3 with unit cell parameters $a = 7.386(1)$ Å and $c = 4.516(1)$ Å. Simpsonite contains only one H atom in the unit cell, which is convenient for this demonstration experiment.

We obtained our data at the N5 instrument located at the National Research Universal (NRU) reactor, Chalk River Laboratories, Canada. The experimental set-up is depicted in an annotated photograph in Fig. 1. Neutrons of wavelength $\lambda = 1.3$ Å with an estimated wavelength resolution of $\Delta\lambda/\lambda \approx 1.5\%$ were obtained from a (113) reflection of a germanium monochromator. The sample-to-detector angular resolution was limited by distance collimation, as defined by a combination of an aperture in neutron-absorbing cadmium masks and the sample size, to approximately 2° vertically and 2° horizontally. The sample orientation (ϕ) and

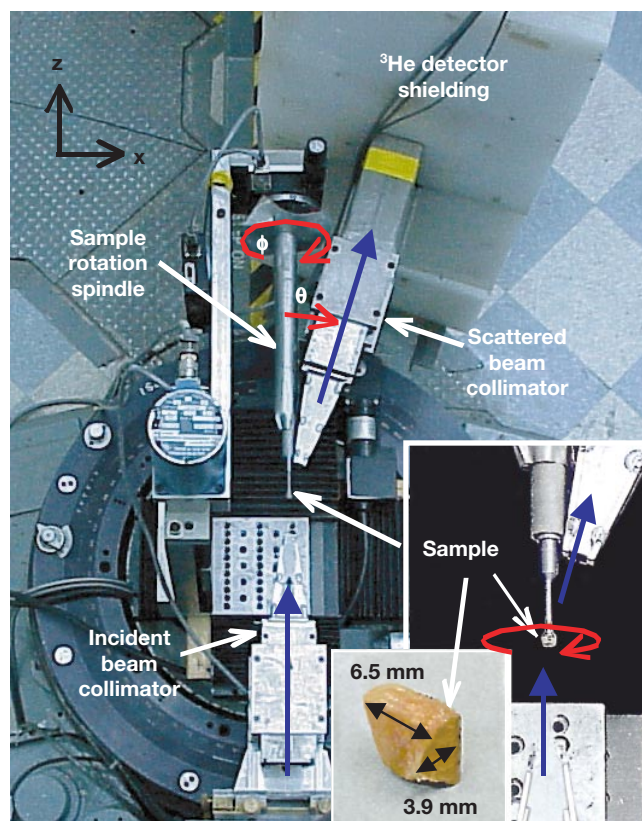


Figure 1 Plan view photograph of the experimental set-up. The incident neutron beam (originating from the bottom of the photograph) is parallel to the sample rotation axis (ϕ). The hologram data was obtained by rotating the sample through 2π radians, in optimal steps for the given detector coverage ($2^\circ \times 2^\circ$), at a given detector angle θ . This was repeated, in 2° steps in θ , for $17^\circ \leq \theta \leq 83^\circ$. Interference between the scattered beam collimator and the sample rotation spindle imposed a lower limit on θ .

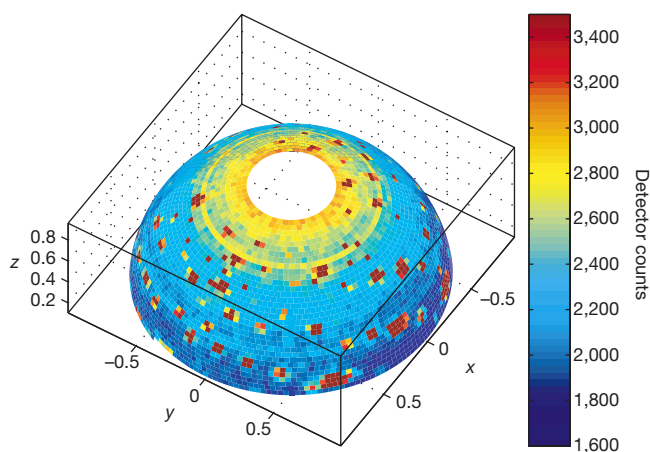


Figure 2 Raw hologram data, plotted in $2^\circ \times 2^\circ$ pixels on a surface of constant scattered wavevector magnitude ($|\mathbf{k}_{\text{out}}| = 4.8 \text{ \AA}^{-1}$). Coordinates are oriented as indicated in Fig. 1, and labelled in units of k_{out} . Detector counts are indicated by colour. Bragg-spot intensities

are truncated at the largest colour value indicated. The thin yellow ring at about 30° from the pole is one of the aluminium lines.

detector angle (θ) were manipulated to optimally tile the scattered wavevector, $\mathbf{k}_{\text{out}}$, over most of a hemisphere of scattering from the sample. Data collection (total scattering and background) over approximately 1.7π steradians in $\mathbf{k}_{\text{out}}$ directions took 10 days of beam time to achieve sufficient statistical quality.

Figure 2 shows the raw intensity data, containing 4,334 solid-angle $2^\circ \times 2^\circ$ pixels limited to $17^\circ \leq \theta \leq 83^\circ$. About 13% of the pixels contain evident, intense Bragg peaks that were judiciously excluded from the data analysis on a statistical basis. The remaining intensity data contains the hologram data plus a θ -dependent background due to the vanadium pedestal and aluminium powder lines (at $\theta = 32.3^\circ, 37.4^\circ, 54.0^\circ$ and 64.3°) due to the aluminium wires used to secure the simpsonite crystal to the vanadium pedestal (Fig. 1). The measurements were repeated without the sample. This measured background, scaled by the measured sample transmission, was subtracted from the raw intensity data. The remaining scattered intensity distribution from the sample, $I(\theta, \phi)$, contains a θ -dependent variation dominated by the H-atom incoherent elastic scattering Debye–Waller factor and an irregular variation in θ and ϕ of the self-attenuation owing to the highly asymmetric shape of the natural simpsonite sample (see smaller inset of Fig. 1). These variations were estimated by a least-squares fit of the background-subtracted intensity to slowly varying, phenomenological functions of the form $D(\theta) = \alpha_1 \exp(\alpha_2(1 - \cos \theta))$ for the Debye–Waller factor, and $S(\theta, \phi) = (\beta_0 + \beta_1\theta + \beta_2\theta^2 + \beta_3\theta^3)(\cos(\phi + \beta_4))$ for the effect of the irregular sample shape (α_i, β_i are fitted parameters). The hologram modulation function is given by:

$$\chi(\theta, \phi) = \frac{I(\theta, \phi)}{D(\theta)S(\theta, \phi)} - 1 \quad (1)$$

From refs 12 and 18, the hologram modulation function for a scattering potential consisting of a strong incoherent scatterer at the origin, surrounded by a coherent scattering length distribution, $b(\mathbf{r})$, of pure s-wave, weak scatterers, is given by (aside from constants):

$$\chi(\mathbf{k}) \approx \int_{\text{all space}} \text{Re} \left(\frac{b(\mathbf{r})}{r} \exp(i(kr - \mathbf{k} \cdot \mathbf{r})) \right) d\mathbf{r} + O(b^2) \quad (2)$$

The conditions for pure s-wave, weak scattering are particularly well satisfied for unpolarized thermal neutron scattering from nuclei. Additionally, for most nuclei, the thermal neutron-scattering lengths are almost purely real and have the same sign¹⁹. As such, we can eliminate the conjugate image from a hologram made with a single thermal neutron wavelength. Thus, the real part of the

scattering potential can be directly reconstructed from the measured hologram modulation function using the following relations:

$$\tilde{b}_{\text{even}}(\mathbf{r}) = \frac{b(\mathbf{r}) + b(-\mathbf{r})}{kr} \cos(kr) \propto \int_{\text{constant } |\mathbf{k}|} \chi(\mathbf{k}) \cos(\mathbf{k} \cdot \mathbf{r}) d\mathbf{k} \quad (3)$$

$$\tilde{b}_{\text{odd}}(\mathbf{r}) = \frac{b(\mathbf{r}) + b(-\mathbf{r})}{kr} \sin(kr) \propto \int_{\text{constant } |\mathbf{k}|} \chi(\mathbf{k}) \sin(\mathbf{k} \cdot \mathbf{r}) d\mathbf{k} \quad (4)$$

For single-wavelength or constant- $|\mathbf{k}|$ data, $b(\mathbf{r})$ cannot be directly determined by inversion of $\tilde{b}_{\text{odd}}(\mathbf{r})$ and $\tilde{b}_{\text{even}}(\mathbf{r})$. However, $\tilde{b}_{\text{odd}}(\mathbf{r}) \sin(kr)$ and $\tilde{b}_{\text{even}}(\mathbf{r}) \cos(kr)$ can be combined by a moving average at the half-wavelength scale to give $b(\mathbf{r})$. Although this procedure permits the elimination of the conjugate image (that is, $b(-\mathbf{r})$) in the reconstruction from single-wavelength data, one consequence is that the atomic positions are pushed outwards from the origin. This is because the neglected intensity of the imaginary part of the reconstructed wave is an exponential, decaying away from the origin. Combining holograms obtained at several different neutron wavelengths will eliminate such artefacts.

Following the above procedure, reconstructions from single-wavelength data are shown in Figs 3 and 4. The reconstructions for this demonstration experiment are not of the highest possible quality for the following reasons: (1) The data is incomplete along θ and leads to a smearing out of the resolution along z of the reconstruction. (2) The counting statistics are just adequate—the statistical uncertainty per pixel is 2 to 3 per cent, whereas the expected hologram modulation, from theoretical considerations and simulations, is of the order of 0.5 per cent. This leads to a noisy reconstruction and limits the range over which the reconstruction is reliable. (3) The hologram reference beam is strongly modulated by the Debye–Waller factor of the source atom. (4) The shape correction does not fully account for the extremely asymmetric shape of the sample. The last two effects tend to produce asymmetric distortions of the real space reconstruction. Nonetheless, the data presented here demonstrate that the holographic *ab initio* reconstruction reproduces the known features of the simpsonite crystal structure (Figs 3 and 4).

The structure of simpsonite contains two layers of oxygen atoms perpendicular to the c axis of the crystal lattice. The first layer, which is closest to the hydrogen atom, has an oxygen atom located directly above (positive direction) the hydrogen atom along the direction of the c axis. This oxygen atom is surrounded by two triangular sublattices of oxygen atoms, slightly below the position of the central oxygen atom. In the second layer, located below (negative direction) the hydrogen atom, the oxygen atoms form triangular lattices about

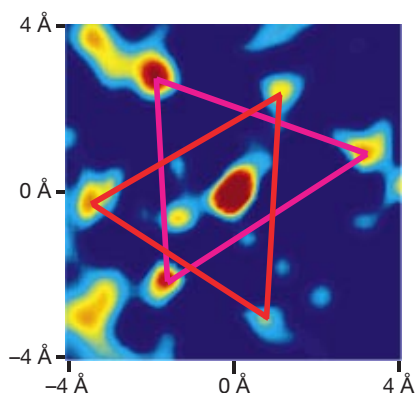


Figure 3 Reconstructed plane located approximately $+0.9 \text{ \AA}$ from the hydrogen atom (the origin). The plane is roughly coplanar to the basal plane of the crystal. The central spot is the oxygen atom located directly above the origin. The triangles are used to indicate the two triplets of oxygen atoms located slightly below this central oxygen atom. The distortion of the triangles is attributed to limitations in the quality of this demonstration data.

the perpendicular projection of the hydrogen atom onto this layer.

Figure 3 shows a reconstructed plane (approximately coplanar to the basal plane) at a distance of $+0.9 \text{ \AA}$ from the incoherent source hydrogen atom (the origin of the reconstruction). There is a systematic distortion of the triangles (see overlays in Fig. 3) owing to the data quality limitations. Nevertheless, all of the expected oxygen atoms, the six of the two triangles and the central atom, are present in the reconstruction. From X-ray data¹⁷ the oxygen–oxygen separation along the side of these triangles is around $4.7 \text{ \AA}$. On the basis of the reconstruction the average of the six oxygen–oxygen separations illustrated in Fig. 3 is around $5.2 \text{ \AA}$ with a deviation from the mean of $\pm 0.3 \text{ \AA}$.

The next closest plane of oxygen atoms, found at a distance of $-1.4 \text{ \AA}$ from the origin, is depicted in Fig. 4. From the reconstruction, it is easily seen that these oxygen atoms form a similarly distorted triangle of nearest neighbours whose average bond length is found to be $3.1 \text{ \AA}$, with a deviation from the mean $\pm 0.5 \text{ \AA}$. X-ray data¹⁷ determine the oxygen bond lengths to be $2.65 \text{ \AA}$. The larger bond length deviation for this plane of oxygen atoms, as compared to those oxygen atoms found at $0.9 \text{ \AA}$, is attributed to a combination of poor vertical resolution and increased distance from the origin.

It should be noted that the distortion of the triangle in Fig. 4 is consistent with the observed distortion of the two triangles in Fig. 3. This systematic distortion provides confidence in the assertion that the observed distortions in the reconstruction are attributable to the effects already described. Consequently, it is expected that an accurate reconstruction can be realized with refinements to the data analysis, and with several data sets at various wavelengths, each approaching 2π steradians in coverage, with improved resolution and statistical quality.

Thus we have shown for the first time, to our knowledge, that atomic-resolution holography can be achieved using thermal neutrons. The technique is based on using the incoherent scattering from atoms of the material as internal, isotropic neutron sources. Subsequent coherent scattering produces the hologram. For this data the very large incoherent scattering cross-section of the hydrogen atom was used to full advantage. This should make the technique of particular interest to studies of polymers and biological materials that are intrinsically rich in hydrogen. Also, because thermal neutrons are low in kinetic energy (of order meV) and carry no charge, they cause little or no damage in most materials, including delicate biological systems^{20,21}. Spence²² discusses in broad terms the implications and potential applications of the holography of atoms. The unique properties of neutrons—they distinguish atoms of similar Z , differentiate isotopes, couple to

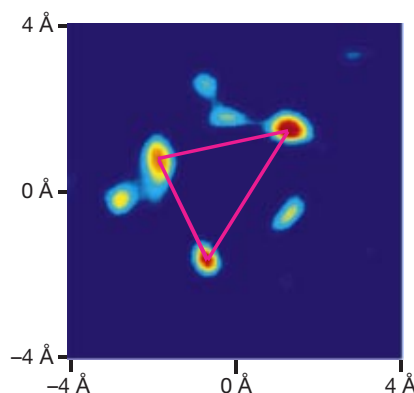


Figure 4 Reconstructed plane located approximately $-1.4 \text{ \AA}$ from the hydrogen atom (the origin). The plane is roughly coplanar to the basal plane of the crystal. The centre of this triangle of nearest-neighbour oxygen atoms is located directly below the origin (defined by the hydrogen atom) in simpsonite. The distortion of the triangle is discussed in the text and the Fig. 3 caption.

magnetic moments, and to first-order coherently scatter equally well from light and heavy atoms—may now be used to address problems that cannot be solved by holography using electromagnetic radiation. In short, owing to the complementarity of neutrons with forms of electromagnetic radiation, thermal neutron atomic-structure holography is expected to become an essential tool for holographic techniques. □

Received 28 June; accepted 6 September 2001.

1. Bragg, W. L. The X-ray microscope. *Nature* **149**, 470 (1942).
2. Gabor, D. A new microscopic principle. *Nature* **161**, 777–778 (1948).
3. Harp, G. R., Saldin, D. K. & Tonner, B. P. Atomic-resolution electron holography in solids with localised sources. *Phys. Rev. Lett.* **65**, 1012–1015 (1990).
4. Zharnikov, M., Weinelt, M., Zebish, P., Stichler, M. & Steinrück, H.-P. First experimental determination of an adsorption site using multiple wave number photoelectron diffraction patterns. *Phys. Rev. Lett.* **73**, 3548–3551 (1994).
5. Sieger, M. T., Roesler, J. M., Lin, D.-S., Miller, T. & Chiang, T.-C. Holography of Ge(111)-(c(2×8) by surface core-level photoemission. *Phys. Rev. Lett.* **73**, 3117–3120 (1994).
6. Orchowski, A., Rau, W. D. & Lichte, H. Electron holography surmounts resolution limit of electron microscopy. *Phys. Rev. Lett.* **74**, 399–402 (1995).
7. Tegze, M. & Faigel, G. Atomic-resolution X-ray holography. *Europhys. Lett.* **16**, 41–46 (1991).
8. Gog, T. *et al.* Multiple-energy X-ray holography: atomic images of hematite (Fe_2O_3). *Phys. Rev. Lett.* **76**, 3132–3135 (1996).
9. Tegze, M. & Faigel, G. X-ray holography with atomic resolution. *Nature* **380**, 49–51 (1991).
10. Tegze, M., Faigel, G., Marchesini, S., Belakhovsky, M. & Chumakov, A. I. Three dimensional imaging of atoms with isotropic $0.5 \text{ \AA}$ resolution. *Phys. Rev. Lett.* **82**, 4847–4850 (1999).
11. Bompadre, S. G., Petersen, T. W. & Sorensen, L. B. Tabletop Bremsstrahlung X-ray holography: making multiwavelength X-ray holograms. *Phys. Rev. Lett.* **83**, 2741–2744 (1999).
12. Cser, L., Krexner, G. & Török, Gy. Atomic-resolution neutron holography. *Europhys. Lett.* **54**(6), 747–752 (2001).
13. Szöke, A. in *Short Wavelength Coherent Radiation: Generation and Applications*, Proc. Topical Meeting (eds Attwood, T. & Boker, J.) 361–467 (AIP Conf. Proc. No. 147, American Institute of Physics, New York, 1986).
14. Bowley, H. Simpsonite (sp. nov.) from Tabba Tabba, western Australia. *J. R. Soc. W. Aust.* **25**, 89–92 (1939).
15. Faigel, G. & Tegze, M. X-ray holography. *Rep. Prog. Phys.* **62**, 355–393 (1999).
16. Taylor, L. E. R. X-ray studies of Simpsonite. *J. R. Soc. W. Aust.* **25**, 93–97 (1939).
17. Ecri, T. S., Černý, P. & Hawthorne, F. C. The crystal chemistry of Simpsonite. *Can. Mineral.* **30**, 663–671 (1992).
18. Barton, J. J. Removing multiple scattering and twin images from holographic images. *Phys. Rev. Lett.* **67**, 3106–3109 (1991).
19. Sears, V. F. Neutron scattering lengths and cross sections. *Neutron News* **3**, 26–37 (1992).
20. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic-resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28**, 171–193 (1995).
21. Weik, M. *et al.* Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc. Natl Acad. Sci. USA* **97**, 623–628 (2000).
22. Spence, J. Holograms of atoms. *Nature* **410**, 1037–1040 (2001).

Acknowledgements

We acknowledge the intellectual contributions of I. P. Swainson and the expert technical assistance of J. H. Fox and L. E. McEwan. We also thank T. S. Ecri for providing the simpsonite crystal.

Correspondence and requests for materials should be addressed to R.B.R. (e-mail: ron.rogge@nrc.ca).

Pressure-induced amorphization and an amorphous–amorphous transition in densified porous silicon

Sudip K. Deb^{*||}, Martin Wilding^{†||}, Maddury Somayazulu^{‡||}
& Paul F. McMillan^{§||}

^{*} Solid State Physics Division, Bhabha Atomic Research Centre, Mumbai 400 085, India

[†] Thermochemistry Facility, University of California at Davis, Davis, California 95616, USA

[‡] HPCAT, Advanced Photon Source, Argonne National Laboratories, Argonne, Illinois 60439, USA

[§] Department of Chemistry, Christopher Ingold Laboratories, University College London, 20 Gordon Street, London WC1H 0AJ, UK, and Davy–Faraday Research Laboratory, Royal Institution of Great Britain, 21 Albemarle Street, London W1X 4BS, UK

^{||} These authors contributed equally to this work

Crystalline and amorphous forms of silicon are the principal materials used for solid-state electronics and photovoltaics technologies. Silicon is therefore a well-studied material, although new structures and properties are still being discovered^{1–4}. Compression of bulk silicon, which is tetrahedrally coordinated at atmospheric pressure, results in a transition to octahedrally coordinated metallic phases⁵. In compressed nanocrystalline Si particles, the initial diamond structure persists to higher pressure than for bulk material, before transforming to high-density crystals⁶. Here we report compression experiments on films of porous Si, which contains nanometre-sized domains of diamond-structured material^{7–9}. At pressures larger than 10 GPa we observed pressure-induced amorphization^{10,11}. Furthermore, we find from Raman spectroscopy measurements that the high-density amorphous form obtained by this process transforms to low-density amorphous silicon upon decompression. This amorphous–amorphous transition is remarkably similar to that reported previously for water^{12,13}, which suggests an underlying transition between a high-density and a low-density liquid phase in supercooled Si (refs 10, 14, 15). The Si melting temperature decreases with increasing pressure, and the crystalline semiconductor melts to a metallic liquid with average coordination ~ 5 (ref. 16).

Porous silicon (π -Si) is of interest because of its luminescence⁸. Samples prepared by electrolytic etching (in aqueous HF solution) consist of nanometre- to micrometre-sized domains of crystalline material separated by voids^{7–9}. The size distribution of crystalline regions within the void walls depends on etching conditions and the starting material. The π -Si used here had a porosity of about 40–50%, and an average crystallite size of ~ 5 nm (refs 7, 17). Analysis of the Raman band profile indicated two distributions of nanocrystallites, with average diameters ~ 3 nm and ~ 7 nm (ref. 17). For high-pressure experiments, 30–50- μ m plates of free-standing π -Si films were placed in a diamond anvil cell with methanol/ethanol or Ar as pressure medium^{17,18}. Synchrotron X-ray diffraction experiments were conducted at the US National Synchrotron Light Source (Brookhaven X-17C). At ambient pressure, π -Si showed red luminescence when irradiated with the Ar⁺ laser used to excite Raman scattering (514.5 or 488 nm excitation). The luminescence shifted to longer wavelength with increased pressure¹⁸, and samples became opaque above ~ 10 GPa.

Upon initial compression to 5 GPa, the broad (111) reflection of π -Si narrowed, indicating annealing of nanocrystalline regions. Above 7–8 GPa, the crystalline diffraction began to disappear, being replaced by broad diffraction characteristic of amorphous

material. By ~ 10 –12 GPa, the crystalline peak was lost (Fig. 1). Raman scattering is more sensitive than X-ray diffraction to remaining traces of crystallinity. Although the crystalline Raman signature weakened above ~ 6 GPa, judging by increased noise in the spectrum, a weak remnant of the crystalline Si peak was retained to 15–17 GPa (Fig. 2). This peak completely disappeared by 19 GPa during normal compression cycles: the last remnant of crystallinity also disappeared on maintaining the sample at 15–16 GPa for greater than 15–30 min. We note that this pressure regime lies above the stability range for diamond-structured Si (ref. 6). Our observations indicate that pressure-induced amorphization (PIA) of crystalline Si has occurred in the sample. PIA has been reported for many materials, including III–V semiconductors—but not for Si, in which recrystallization is usually rapid even at ambient temperature^{10,11}. We found no evidence for appearance of dense crystalline phases of Si on compression in excess of 10 GPa, unlike the behaviour observed for bulk material or surface-terminated nanocrystalline particles^{5,6}. We believe that this is due to the nature of the sample, and transformation mechanisms to the high density phases.

At the highest pressures, the Raman spectrum of high-density amorphous (HDA) silicon consisted of a weak, broad band between 200 and 400 cm^{-1} (Fig. 2). This spectrum differs markedly from the Raman signature of normal, tetrahedral, amorphous silicon (a-Si: the low-density amorphous, LDA, form), which is a broadened version of the vibrational density of states of crystalline diamond-structured silicon¹⁹. Silicon with the β -Sn structure has an evenly distributed spectrum of vibrational modes extending to $\sim 350 \text{ cm}^{-1}$ at 10–20 GPa (ref. 20), which resembles the broad Raman band observed for the high-pressure a-Si form reported here. We conclude that PIA of π -Si has produced an HDA form of the element, which may structurally resemble the β -Sn crystalline form, with Si

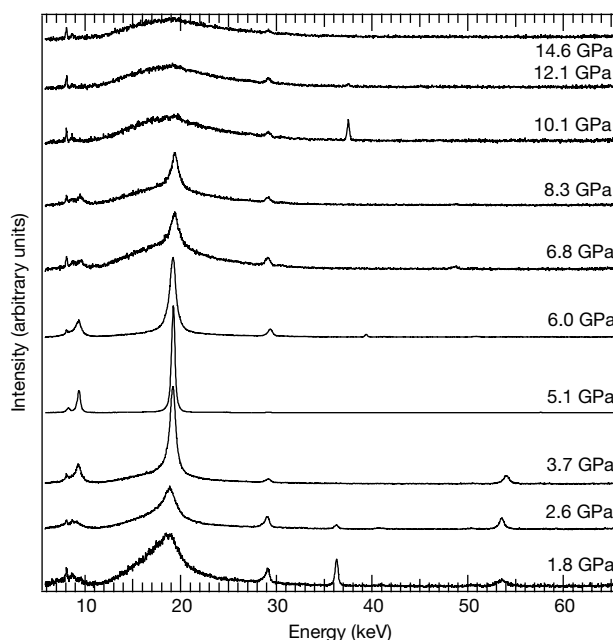


Figure 1 Energy-dispersive X-ray diffraction from porous silicon at high pressure. Data were obtained at $2\theta = 12.002^\circ$. Samples of porous silicon (π -Si) were loaded in Mao–Bell diamond anvil cells with 4:1 methanol/ethanol or Ar as pressure medium, along with ruby chips for pressure measurement. A background function due to the beam profile (including Compton scattering from the diamonds) was subtracted from the data. The 1-atm spectrum (not shown) is dominated by the Si (111) reflection ($d_{111} = 3.135 \text{ \AA}/18.9 \text{ keV}$ for bulk Si), broadened by sample nanocrystallinity. The average particle dimension is 5 nm. Additional peaks near 29 and 52 keV are due to the Re gasket and ruby used for pressure calibration.

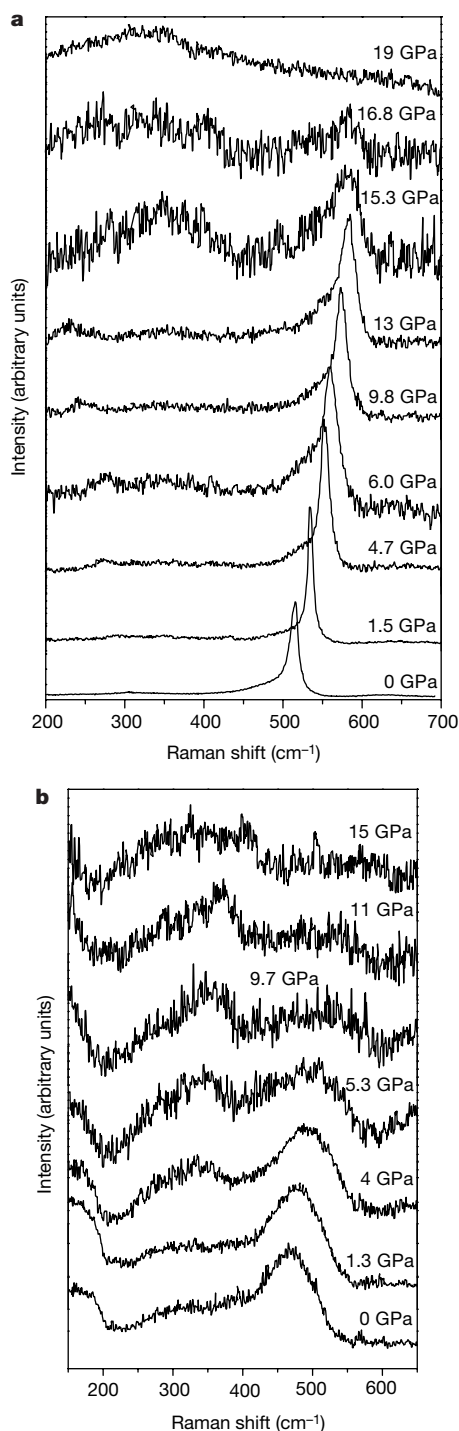


Figure 2 Raman spectra of porous silicon during one compression/decompression cycle. The ambient pressure spectrum (shown in **a**) is typical of nanocrystalline Si, dominated by an asymmetric peak at $\sim 520\text{ cm}^{-1}$ due to zone centre phonons. Analysis of the band profile indicates two distributions of nanocrystalline domains, with average diameters $\sim 3\text{ nm}$ and $\sim 7\text{ nm}$. **a**, Changes in Raman spectra during compression. The π -Si film gradually lost transparency above 7–8 GPa, and by 13–15 GPa the sample was completely black. The sample remained opaque during decompression. **b**, During decompression, only the weak, broad amorphous feature at $200\text{--}400\text{ cm}^{-1}$ was visible initially. Below $\sim 8\text{--}9\text{ GPa}$, this feature was joined by a second broad band at $400\text{--}500\text{ cm}^{-1}$, indicative of the tetrahedrally coordinated low-density amorphous (LDA) form of a-Si. Below 4 GPa, the signal-to-noise ratio of the spectra improved dramatically, and the spectral features in the $250\text{--}400\text{ cm}^{-1}$ region became considerably reduced in intensity, signalling the complete return of the amorphous material from the high-density amorphous (HDA; metallic amorphous) state to the normal tetrahedral amorphous semiconductor (LDA form).

coordination of greater than 4. Upon decompression, only the Raman signature of HDA Si was present to $\sim 10\text{ GPa}$, at which point a broad amorphous band at $\sim 470\text{ cm}^{-1}$ grew in the spectrum, indicating the reappearance of tetrahedrally coordinated LDA Si (Fig. 2).

Pressure-induced amorphization occurs when the metastable extension of the melting curve (dT_m/dP) of a low-density crystal phase is intercepted during compression¹⁰ (here T_m is the melting temperature, and P is pressure). Extrapolation of the negative melting slope of Si leads to prediction of PIA at $\sim 14\text{ GPa}$ at temperature $T = 300\text{ K}$, consistent with our observations (Fig. 3). Systems showing PIA are also candidates for density-driven

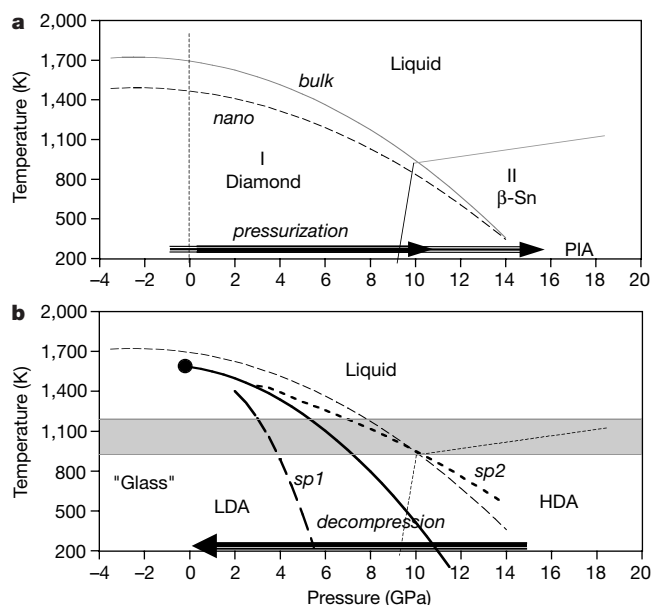


Figure 3 Calculated temperature–pressure (T – P) diagrams of Si in stable and metastable crystalline and amorphous states. **a**, Diagram showing the negatively curved melting slope for bulk Si (diamond structure), and its metastable extrapolation into the phase field of the metallic β -Sn (Si-II) phase, as well as the depressed melting line for nanocrystalline Si (calculated for 5-nm particles^{29,30}). PIA occurs experimentally at 10–15 GPa during metastable compression at 300 K. The metastable extension of the melting curve intercepts the pressurization line at $P \approx 14\text{ GPa}$, where the final disappearance of the Raman peak of crystalline Si is observed in the experiments. **b**, Calculated phase relations in the supercooled liquid, using the bond excitation model to describe the liquid behaviour²¹. Model thermodynamic parameters used were $\Delta H = 40\text{ kJ mol}^{-1}$, $\Delta S = 24.9\text{ J mol}^{-1}\text{ K}^{-1}$ and $\Delta V = -0.91\text{ cm}^3$, expressed as differences between low- and high-density states, which are consistent with available thermodynamic data and computer simulations of liquid Si (refs 10, 24, 25). A value of the non-ideal interaction (mixing) parameter $W = 26,604\text{ J mol}^{-1}$ successfully reproduced the measured negative melting slope, with a predicted maximum at approximately -3 GPa . The set of values also gave a calculated critical point for the liquid–liquid transition at approximately -0.5 GPa and $T \approx 1,600\text{ K}$. A line of calculated first-order liquid–liquid phase transitions extends to higher P and lower T from the critical point. Density measurements on supercooled liquid Si show no observed anomaly to $T \approx 1,350\text{ K}$ (ref. 23): this may indicate that the actual transition temperature is very sensitive to the model parameters in the region of the critical point, or that the first order nature of the HDL–LDL transition results in slow kinetics, or that there is only a small density change at the transition temperature (that is, the transition is largely driven by the entropy difference between HDL and LDL phases at high T). The stippled band shows the approximate location of the glass transformation estimated for HDL and LDL forms of silicon (Fig. 4). In our experiments, HDA Si is formed directly below T_g by PIA. During decompression, the HDA–LDA transition is encountered at 300 K at 10–11 GPa. The $400\text{--}500\text{ cm}^{-1}$ band, which is the signature of the LDA form, begins to appear in Raman spectra below 10 GPa. The one-phase spinodal (sp1) occurs at $\sim 5\text{ GPa}$ on decompression. We can only distinguish the signature of LDA Si in Raman spectra below this pressure (Fig. 2).

liquid–liquid phase transitions^{10,14,15}. Two-state models used to describe liquids that exhibit melting-curve maxima or negative melting slopes consider mixing of species with low and high coordination¹⁶. Application of regular solution thermodynamics to the mixing reveals a critical point, below which first-order transitions occur between low- and high-density liquid (LDL and HDL) phases: that is, the liquid-state behaviour passes from two-state to two-phase¹⁰. A recent model considers bond-excitation processes rather than distinct structural species to describe the liquid structure²¹. We studied the liquid–liquid transition behaviour for Si using a bond-excitation model constrained by available thermodynamic data^{5,10,22}. We predict a first-order transition between an HDL form with Si in high coordination, resembling metallic crystalline phases, and a tetrahedrally coordinated LDL semiconductor form occurring at low pressure and temperature (Fig. 3). At 300 K, the extrapolated transition occurs at ~ 10 GPa, where we observed initial reappearance of the Raman band of LDA Si (Fig. 2). The calculated one-phase spinodal occurs at ~ 4 GPa. Below this pressure, only the LDA amorph was observed in our spectra (Fig. 2).

The HDL and LDL forms of liquid Si should have different rheologies. We calculated viscosities of liquid Si from the calculated entropy²², using different mixing parameter values (W) corresponding to (1) HDL behaviour throughout the temperature range and (2) resulting in an HDL–LDL critical point transition (Fig. 4). Smaller W (13,350 J mol^{−1}) corresponding to the behaviour of HDL Si throughout the entire temperature range yields a calculated viscosity (η) consistent with measurements for the stable liquid²³. The point at which the extrapolated curve reaches $\eta = 10^{12}$ Pa s ($T = 1,070$ K) corresponds to the suggested glass transition (T_g) for liquid Si (ref. 24). The $\log \eta - (1/T)$ relation is strongly curved, as expected for a “fragile” liquid²⁵. A larger value of W (26,600 J mol^{−1}) results in a predicted critical point at $T_c \approx 1,600$ K and -0.5 GPa, and a first-order phase transition occurring between HDL and LDL below T_c at 1 atm (Fig. 3). Below the HDL–LDL transition, the $\log \eta - (1/T)$ relation is less strongly curved, indicating that the proposed LDL silicon is “stronger” (less fragile) than the HDL form. We predict that the T_g for LDA Si lies at a higher temperature ($T = 1,280$ K for $\eta = 10^{12}$ Pa s) than that for the HDA form. For rapid heating rates, of the order of 10^3 – 10^6 K s^{−1}, the expected LDA T_g would rise to 1,300–1,400 K. Flash heating experiments on a-Si show a transition within the amorphous state at $T = 1,345$ K (refs 26, 27), which has been interpreted as an LDA–HDL transition. It could also be due to the LDA glass transition.

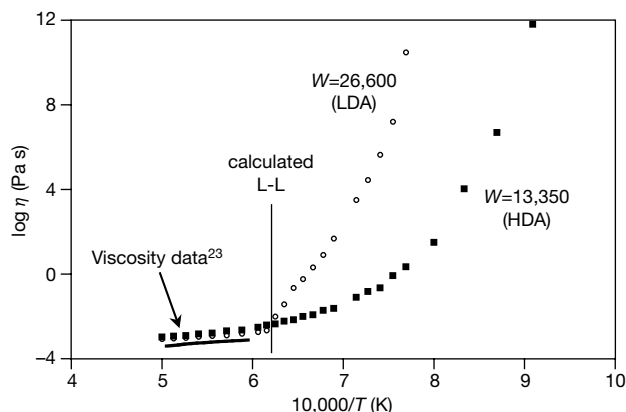


Figure 4 Calculated viscosities of hypothetical HDL and LDL silicon liquids. The configurational entropies were obtained from the bond-lattice excitation model for HDL and LDL phases²¹ and were used to predict the viscosity (η) via the Adam–Gibbs relation²². Two values of the interaction parameter W were used to model the behaviour of silicon liquids over the entire temperature range. See text for details. Both values of W yield calculated $\eta - T$ relations that are consistent with experiment above $\sim 1,650$ K (ref. 23).

Our observations suggest an underlying HDL–LDL phase transition in supercooled liquid Si, between a metallic liquid phase with Si in high coordination and a tetrahedrally coordinated amorphous semiconductor. The transformation is observed here below the glass transition for both liquids. The behaviour is directly linked to the observed negative melting slope of diamond-structured Si and to our observation of PIA in the crystalline material, and it provides further support for the occurrence in simple systems of density- and entropy-driven transitions. These are a newly recognized but apparently general phenomenon in the physical chemistry of liquids^{10,14,15,28}. □

Received 8 June; accepted 3 October 2001.

- Piltz, R. O. *et al.* Structure and properties of silicon XII: A complex tetrahedrally bonded phase. *Phys. Rev. B* **52**, 4072–4085 (1995).
- Ramachandran, G. K. *et al.* High-pressure phase transformation of the silicon clathrate Si₁₃₆. *J. Phys. Condens. Matter* **12**, 4013–4020 (2000).
- Kamalkaran, R., Singh, A. K. & Srivastava, O. N. Formation and characterization of nanoparticle-bearing threads of silicon, germanium and tin. *J. Phys. Condens. Matter* **12**, 2681–2689 (2000).
- Blanco, A. *et al.* Large-scale synthesis of a silicon photonic crystal with a complete three-dimensional bandgap near 1.5 micrometres. *Nature* **405**, 437–440 (2000).
- Liu, L.-G. & Bassett, W. A. *Elements, Oxides and Silicates: High Pressure Phases with Implications for the Earth's Interior* 37 (Clarendon, Oxford, 1987).
- Tolbert, S. H., Herhold, A. B., Brus, L. E. & Alivisatos, A. P. Pressure-induced structural transformations in Si nanocrystals: surface and shape effects. *Phys. Rev. Lett.* **76**, 4384–4387 (1996).
- Xie, Y. H. *et al.* Absorption and luminescence studies of free-standing porous silicon fibres. *Phys. Rev. B* **49**, 5386–5397 (1994).
- Collins, R. T., Fauchet, P. M. & Tischler, M. A. Porous silicon: from luminescence to LEDs. *Phys. Today* January 24–31 (1997).
- Bellet, D. & Dolino, G. X-ray diffraction studies of porous silicon. *Thin Solid Films* **276**, 1–6 (1996).
- Ponyatovsky, E. G. & Barkalov, O. I. Pressure-induced amorphous phases. *Mater. Sci. Rep.* **8**, 147–191 (1992).
- Sharma, S. M. & Sikka, S. K. Pressure-induced amorphization of materials. *Prog. Mater. Sci.* **40**, 1–77 (1996).
- Mishima, O., Calvert, L. D. & Whalley, E. “Melting” ice at 77 K and 10 kbar: a new method of making amorphous solids. *Nature* **310**, 393–395 (1984).
- Mishima, O., Calvert, L. D. & Whalley, E. An apparently first-order transition between two amorphous phases of ice induced by pressure. *Nature* **314**, 76–78 (1985).
- Poole, P. H., Grande, T., Angell, C. A. & McMillan, P. F. Polymorphic phase transitions in liquids and glasses. *Science* **275**, 322–323 (1997).
- Mishima, O. & Stanley, H. E. The relationship between liquid, supercooled and glassy water. *Nature* **396**, 329–335 (1998).
- Ansell, S. A., Krishnan, S., Felten, J. J. & Price, D. L. Structure of supercooled liquid silicon. *J. Phys. Condens. Matter* **10**, L73–L78 (1998).
- Deb, S. K. in *Adv. High Pressure Sci. Technol. Proc. IV NCHST, IGCAR, Kalpakkam, India* (eds Yousuf, M., Subramanian, N. & Govind, R. K.) 147–152 (Universities Press (India), Hyderabad, 1997).
- Papadimitriou, D., Raptis, Y. S. & Nassiopoulou, A. G. High-pressure studies of photoluminescence in porous silicon. *Phys. Rev. B* **58**, 14089–14093 (1998).
- Brodsky, M. H. & Lurio, A. Infrared vibrational spectra of amorphous Si and Ge. *Phys. Rev. B* **9**, 1646–1651 (1974).
- Gaál-Nagy, K. *et al.* Temperature and dynamical effects on the high-pressure cubic-diamond β -tin phase transition in Si and Ge. *Phys. Status Solidi B* **211**, 275–280 (1999).
- Moynihan, C. T. & Angell, C. A. Bond-lattice or excitation model analysis of the configurational entropy of molecular liquids. *J. Non-Cryst. Solids* **274**, 131–138 (2000).
- Adam, G. & Gibbs, J. H. On the temperature dependence of cooperative relaxation properties in glass forming liquids. *J. Chem. Phys.* **43**, 139–146 (1965).
- Rhim, W. K. & Ohsaka, K. Thermophysical properties measurement of molten silicon by high-temperature electrostatic levitator: density, volume expansion, specific heat capacity, emissivity, surface tension and viscosity. *J. Cryst. Growth* **208**, 313–321 (2000).
- Angell, C. A., Borick, S. A. & Grabow, M. Glass transitions and first-order liquid-metal-semiconductor transitions in 4–5–6 covalent systems. *J. Non-Cryst. Solids* **205–207**, 463–471 (1999).
- Moynihan, C. T., Angell, C. A. & Hemmati, M. Strong and superstrong liquids and an approach to the perfect glass state via phase transition. *J. Non-Cryst. Solids* **274**, 319–331 (2000).
- Thompson, M. O. *et al.* Melting temperature and explosive crystallisation of amorphous silicon during pulsed laser irradiation. *Phys. Rev. Lett.* **52**, 2360–2363 (1984).
- Donovan, E. P., Spaepen, F., Turnbull, D., Poate, J. M. & Jacobsen, D. C. Calorimetric studies of crystallisation and relaxation of amorphous Si and Ge prepared by ion implantation. *J. Appl. Phys.* **57**, 1795–1804 (1985).
- Katayama, Y. *et al.* A first-order liquid-liquid phase transition in phosphorus. *Nature* **403**, 170–173 (2000).
- Goldstein, A. N., Echer, C. M. & Alivisatos, A. P. Melting in semiconductor nanocrystals. *Science* **256**, 1425–1428 (1992).
- Lai, S. L., Guo, J. Y., Petrova, V., Ramanath, G. & Allen, L. H. Size-dependent melting properties of small tin particles: nanocalorimetric measurements. *Phys. Rev. Lett.* **77**, 99–102 (1996).

Acknowledgements

P.F.M. is a Wolfson–Royal Society Research Merit Award holder.

Correspondence and requests for materials should be addressed to P.F.M. (e-mail: paulm@ri.ac.uk).

Real-time spectroscopy of transition states in bacteriorhodopsin during retinal isomerization

Takayoshi Kobayashi*, Takashi Saito* & Hiroyuki Ohtani†

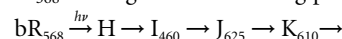
* Department of Physics, Graduate School of Science, University of Tokyo, Hongo 7-3-1, Bunkyo, Tokyo 113-0033, Japan

† Department of Biomolecular Engineering, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259, Nagatsuta, Yokohama 226-8501, Japan

Real-time investigations of the rearrangement of bonds during chemical transformations require femtosecond temporal resolution, so that the atomic vibrations within the reacting molecules can be observed. Following the development of lasers capable of emitting ultrashort laser flashes on this timescale, chemical reactions involving relatively simple molecules have been monitored in detail, revealing the transient existence of intermediate species as reactants are transformed into products^{1–3}. Here we report the direct observation of nuclear motion in a complex biological system, the retinal chromophore of bacteriorhodopsin (bR₅₆₈)⁴, as it undergoes the *trans*–*cis* photoisomerization that is fundamental to the vision process. By using visible-light pulses of less than 5 femtosecond in duration^{5,6}, we are able to monitor changes in the vibrational spectra of the transition state and thus show that despite photoexcitation of the anti-bonding molecular orbital involved, isomerization does not occur instantly, but involves transient formation of a so-called ‘tumbling state’. Our observations thus agree with growing experimental^{7–14} and *ab initio* evidence^{15,16} for a three-state photoisomerization

model^{8–10,17} and firmly discount the initially suggested two-state model^{18–20} for this process.

Using pulses shorter than 5 fs, we have succeeded in reaching the fundamental limit in resolving nuclear motion in a complex biological system. bR₅₆₈ undergoes the following photochemical cycle:



Here, H denotes the Franck–Condon excited state of bR₅₆₈ with an all-*trans* retinal Schiff base. The subscripts represent the wavelengths (in nm) of absorption maxima of the species. Several vibrational spectroscopy studies using ultrashort pulses^{7,21–24} have shown that H and I₄₆₀ are excited-state species with a chromophore of an all-*trans* conformation, and that K₆₁₀ is a ground-state species with the chromophore already converted to the 13-*cis* conformation. But the primary molecular processes between the H and K₆₁₀ states have been controversial, with two models discussed for the photoisomerization process: the two-state^{18–20} model and the three-state model^{8–10,17} (Fig. 1). However, recent experiments^{7–14} and *ab initio* calculations^{15,16} favour the three-state model, with changes in the molecule’s structure assumed to precede the *trans*–*cis* isomerization^{8–16}.

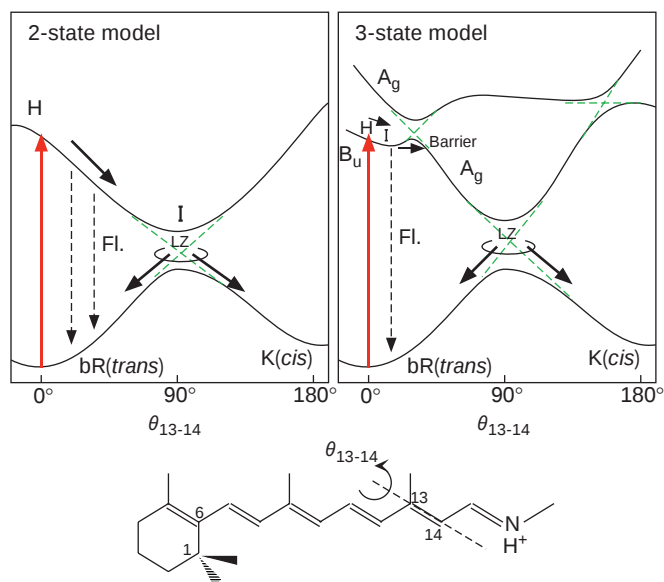


Figure 1 Schematic potential energy curves of two-state and three-state models along the torsional angle of the C₁₃ = C₁₄ bond of the protonated retinal Schiff base of bR₅₆₈. In the two-state model, the wave packet ‘slides down’ the barrier-free smooth potential surface along the C₁₃ = C₁₄ torsion coordinate from the Franck–Condon (H) state to the 90° rotated conformer (I) within 200 fs, and a temporal change in the emission spectrum (dynamic Stokes shift) is expected corresponding to the H→I₄₆₀ process. In the three-state model, the skeletally deformed state (I) is converted to a tumbling state with a time constant of 200 fs and then the torsional angle distribution becomes narrow. In both models K is in a cisoid conformation.

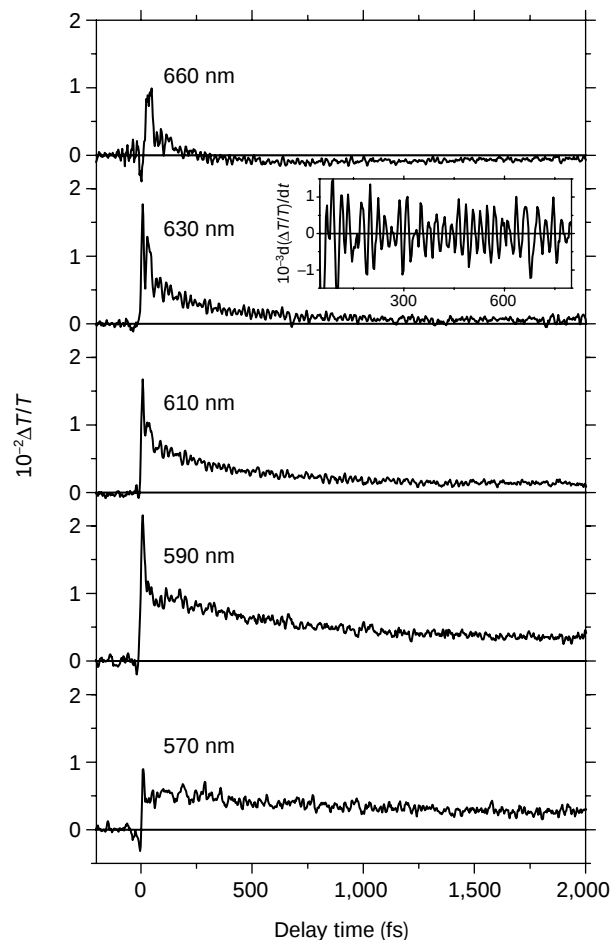


Figure 2 Time-dependence of the transmittance change ($\Delta T/T$) following sub-5-fs pulse excitation of a suspension of the light-adapted purple membrane of *Halobacterium salinarum* containing bR₅₆₈ at room temperature. Sub-5-fs pump and probe pulses (520–750 nm) were generated by non-collinear optical parametric amplification^{5,6}. The positive $\Delta T/T$ observed at wavelengths shorter than 630 nm is due to the bleaching of bR₅₆₈. At wavelengths longer than 660 nm, a negative signal due to the absorption of I (S_n–S₁ transition) appeared at 250 fs. The inset shows an enlarged trace monitored at 630 nm.

In order to elucidate the isomerization process in more detail, we measured the time-resolved transmittance change ($\Delta T/T$), which we call the real-time signal, of bR_{568} using sub-5-fs pulses as both pump and probe at room temperature. Figure 2 shows the real-time signals of bR_{568} at several wavelengths. The measured $\Delta T/T$ is modulated by molecular vibrations, which correspond to Raman-active modes^{25,26}. The Fourier transformation of $\Delta T/T$ around an arbitrary delay time t_d gives the amplitudes and frequencies of the molecular vibrations at t_d . Figure 3 shows spectrograms calculated for real-time signals probed at 610 nm. As shown in the spectrograms, instantaneous frequencies are modulated at a period of about 200 fs. This can be seen from the time-integrated Fourier power spectrum of the vibrationally induced transmittance modulation. Figure 4 shows the results of peak detection using the previously measured Raman frequencies^{23,27,28} for the 630-nm probe data. We conclude that the temporal shifts in frequency are due to wave-packet motion of the excited state rather than the ground state for the following reasons. (1) The spectrogram calculated for the real-time signal at 660 nm due to excited-state absorption (shown in Fig. 2) has the same features as those at other wavelengths. (2) The pulse width (sub-5-fs) is much shorter than the observed molecular vibration periods, and is too short to enhance the molecular vibration in the ground electronic state^{25,26}. (3) The decay of the molecular vibration amplitude is not longer than that of $\Delta T/T$.

The key feature of the spectrogram is the frequency of the ethylene-like symmetric C=C stretching modes in the 1,500–1,550 cm^{-1} region reflecting π -bond order. Interestingly, the peak frequency is modulated between 1,500 and 1,550 cm^{-1} . A similar

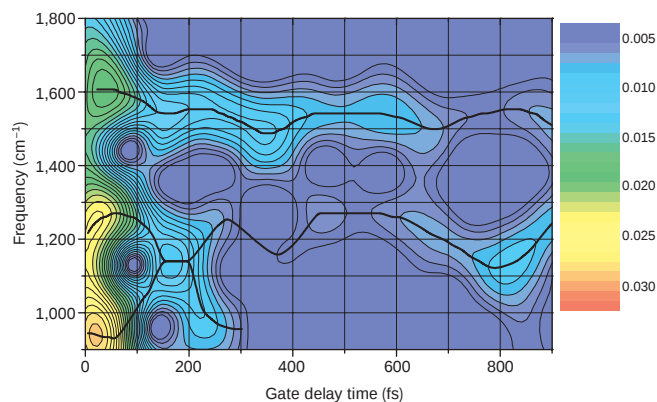


Figure 3 Spectrogram calculated for the trace at 610 nm. The calculations used the Hanning window function $W(t - t_d) = 1/2 + (1/2)\cos[\pi(t - t_d)/2\Delta t]$ with a 75-fs half-width at half-maximum (HWHM) of Δt and a peak at t_d (window delay time). The Fourier transformation of $\Delta T/T$ kinetics gives frequency information that corresponds to Raman frequency. The Fourier amplitude increases from blue to red (false colours). Black lines show the peaks of several modes strongly coupled to photoexcitation. The spectrogram shows the appearance of all the significant vibrations^{23,27,28} in the characteristic frequency ranges 1,500–1,650 cm^{-1} (C=C and C=N stretching modes), 1,150–1,250 cm^{-1} (in-plane C=C–H bending coupling with the C=C stretching modes) and 900–1,000 cm^{-1} (the hydrogen-out-of plane (HOOP) mode). The HOOP mode is inactive when the molecular structure is highly planar, and is enhanced by distortion of the molecular plane. The increase in the HOOP signal intensity just after excitation (30 fs) shows the distortion. The changes in the frequencies and amplitudes of HOOP and C=C–H in-plane-wagging can be seen following the progress of the twisting around the $\text{C}_{13}=\text{C}_{14}$ bond. Their frequencies cannot be discriminated at 150–200 fs because the in-plane and out-of-plane modes cannot be well defined owing to the non-planarity of the molecule. After that, these vibrations recover and then their intensities are reduced substantially because of the broad distributions of frequencies for the 200–600 fs region. The intensities partially recover when the frequency distributions become narrow owing to thermalization in the 700–900 fs region. This shows that isomerization from transoid to cisoid conformation followed by thermalization is over at this time region.

modulation of instantaneous frequencies has been observed for polydiacetylene (PDA) in previous work using the same sub-5-fs visible pulses^{29,30}. In PDA, the frequencies of C=C and C–C stretching are modulated at the vibrational period of the C=C–C bending mode (145 fs) for about 2 ps. The modulation frequency in bR_{568} was calculated to be 160 cm^{-1} from the Fourier transform calculation of the Fourier power spectrum, which corresponds to the 200-fs oscillation period of twisting around the C=C bond. To our knowledge, this is the first real-time observation of the frequency modulation of C=C stretching associated with the torsional motion relevant to the *trans*–*cis* isomerization. The frequency modulation induces the generation of side bands, but conventional Raman spectroscopy cannot discriminate the bands from modes existing accidentally nearby.

We now examine the contribution of the 200-fs torsional motion to the coupling of other vibrational modes by describing four events that occur on different timescales. The first event occurs within 50–100 fs. The wave-packet reaches the flat potential region (I_{460} with a B_u -like electronic configuration) from H. Periodic out-of-plane distortion also begins. An intense C=N stretching band from the protonated Schiff base ($\text{C}_{15}=\text{NH}^+$, 1,620–1,630 cm^{-1}) appears within 50 fs. Both the intensity and frequency of the band dramatically decrease within 100 fs. The fast dumping causes spectral broadening; although conventional methods cannot distinguish the homogeneous and inhomogeneous broadening, the method we describe here—of observing the time-dependent molecular vibration—can distinguish the two, by detecting the recurrence in the

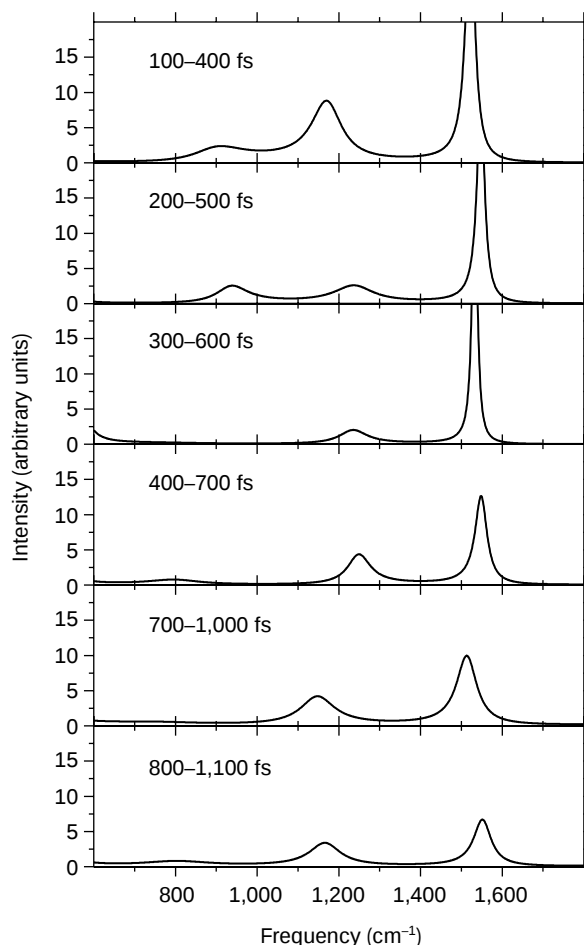


Figure 4 Fourier power spectrum of instantaneous frequencies appearing in the trace of transmittance change probed at 630 nm. The spectra were obtained by the maximum entropy method for the 300-fs rectangular gated signals.

case of inhomogeneous broadening. The fast frequency lowering indicates the effect of the torsional motion. The in-plane C=C-H bending modes coupled with C-C stretching modes appear in the 1,150–1,300 cm^{-1} region, which is called the fingerprint region, because it is very sensitive to the chromophore conformation. The skeletal vibrations around 1,250 cm^{-1} are activated in the first 100 fs as predicted theoretically^{15,16}. The hydrogen-out-of-plane (HOOP) modes appear just after the excitation in the 800–1,050 cm^{-1} region. The intensity of the HOOP modes is sensitive to the tilting angle around the C=C bond. We expect that the HOOP intensity at first increases with the torsional angle, and then decreases or even disappears as the angle approaches 90°.

The second event is seen in the 100–200 fs region, and involves splitting of the band centred around 1,250 cm^{-1} into two peaks about 100 fs after excitation of the chromophore. The frequency of one peak gradually increases, reaching 1,520–1,540 cm^{-1} at about 150 fs, while the frequency of the other peak gradually decreases, reaching a value as low as 1,140 cm^{-1} at 170 fs after excitation. The temporal shift of the frequency of the latter mode (C=C-H in-plane bending) is correlated with that of the HOOP band. These frequencies are modulated with the 200-fs period of the C=C=C torsion. The peak positions of the HOOP mode and the in-plane C=C-H bending modes are well separated in the region below 100 fs, but not in the 150–200 fs region. These mode frequencies approach each other and merge into a single peak, as there is neither an in-plane nor an out-of-plane mode in a distorted configuration. The two modes reappear in the region slightly longer than 250 fs, where planarity is recovered. This suggests that the *trans*→13-*cis* isomerization occurred within less than 200 fs according to the two-state model. However, the product K₆₁₀ has not yet been formed. For the $t_d > 200$ fs region, the intensities of the C=C-H bending and HOOP modes become very weak at all the probe photon energies studied (610–680 nm, data not shown). Therefore, the coincident increase and decrease of the frequencies of the HOOP and in-plane C=C-H bending modes strongly indicate that the torsional

motion modifies the frequencies of both the out-of-plane and in-plane bending-mode frequencies with a period of 200 fs.

The third event is the formation of the transition state in the 200–600 fs region. C=N stretching, C=C-H bending, and HOOP modes are also silent in this region. The weak intensity is partly due to the broad distribution of molecular conformations covered by the wave packet. Energy randomization occurs during the fast wave-packet motion along the potential curves of the excited B_u-like and A_g-like states, and during the Landau–Zener (LZ) tunnelling in between B_u and A_g. However, the isomerization in the C₁₃=C₁₄ bond has not occurred at this stage. The skeletal change precedes the isomerization. We call the molecular state in the 200–600 fs region a ‘tumbling state’. In this state, the intensity reduction of C=C stretching is not as drastic as those of the HOOP and in-plane bending, as the number of the C=C stretching modes is larger than that of the C=C-H configuration, and the C=C stretching is less sensitive to the torsional angle than the HOOP and in-plane bending modes. These features can be explained by the torsional motion of the two moieties of the retinal Schiff base separated by the C₁₃=C₁₄ double bond just after excitation, as shown in Fig. 5. The torsional motion and the broadening due to the inhomogeneity of the torsional angles and the rapid change in frequencies induce the reduction of the intensities of the corresponding vibrational bands.

The final event occurs 700 fs later. At 700–900 fs, the intensities of the in-plane C=C-H bending and C=C stretching modes (1,150–1,250 cm^{-1}) are partially recovered, owing to the narrowing of the torsional angle distribution in the quasi-thermal state after relaxing to the bottom of the potential curve of the A_g-like excited state along the torsional angle coordinate. After this relaxation to the ground state, the conformational change takes place again through LZ tunnelling to the 13-*cis* conformation, that is, K₆₁₀, with 55% efficiency¹⁴. This verifies that the oscillating signal is not caused by the molecular vibration in the ground electronic state, but by vibration in the excited states. The gate delay time of 700–800 fs corresponds to the period of the population thermalized before it is converted to the K intermediate. □

Received 14 May; accepted 21 September 2001.

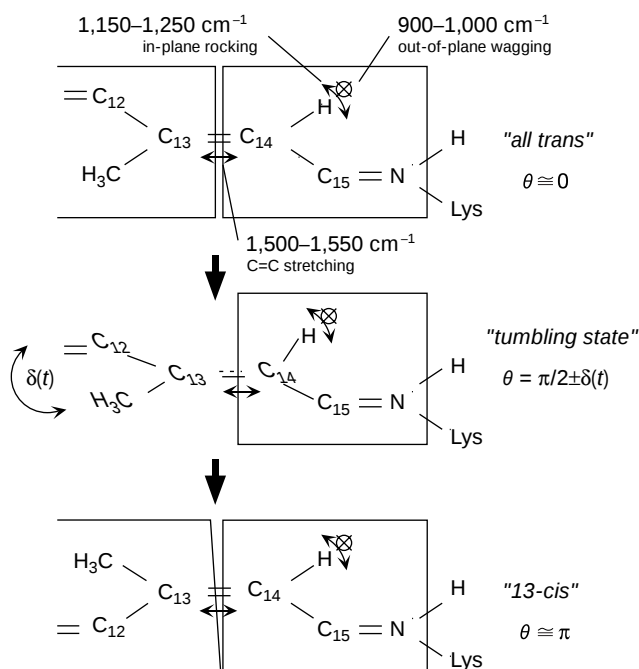


Figure 5 A mechanism of the *trans*–*cis* photoisomerization of the retinal chromophore. After photoexcitation of bR₅₆₈ to the Franck–Condon state (H), two moieties are rotated with respect to each other to a skeletally deformed state (I₄₆₀; top) and then to a tumbling state (middle) and then to a *cis* conformation (K₆₁₀; bottom).

- Polanyi, J. C. & Zewail, A. H. Direct observation of the transition state. *Acc. Chem. Res.* **28**, 119–132 (1995).
- Rose, T. S., Rosker, M. J. & Zewail, A. H. Femtosecond real-time observation of wave packet oscillations (resonance) in dissociation reactions. *J. Chem. Phys.* **88**, 6672–6673 (1988).
- Rose, T. S., Rosker, M. J. & Zewail, A. H. Femtosecond real-time probing of reactions. IV. The reactions of alkali halides. *J. Chem. Phys.* **91**, 7415–7436 (1989).
- Oesterheld, D. & Stoekenius, W. in *Methods of Enzymology, Biomembranes Part A*, Vol. 31 (eds Freischer, S. and Packer, L.) 667–678 (Academic, New York, 1974).
- Shirakawa, A., Sakane, I. & Kobayashi, T. Pulse-front-matched optical parametric amplification for sub-10-fs pulse generation tunable in the visible and near infrared. *Opt. Lett.* **23**, 1292–1294 (1998).
- Shirakawa, A., Sakane, I., Takasaka, M. & Kobayashi, T. Sub-5-fs visible pulse generation by pulse-front-matched noncollinear optical parametric amplification. *Appl. Phys. Lett.* **74**, 2268–2270 (1999).
- Atkinson, G. H., Ujj, L. & Zhou, Y. Vibrational spectrum of the J-625 intermediate in the room temperature bacteriorhodopsin photocycle. *J. Phys. Chem. A* **104**, 4130–4139 (2000).
- Hasson, K. C., Gai, F. & Anfinsen, P. A. The photoisomerization of retinal in bacteriorhodopsin: experimental evidence for a three-state model. *Proc. Natl Acad. Sci. USA* **93**, 15124–15129 (1996).
- Haran, G. *et al.* Excited state dynamics of bacteriorhodopsin revealed by transient stimulated emission spectra. *Chem. Phys. Lett.* **261**, 389–395 (1996).
- Gai, F., Hasson, K. C., McDonald, J. C. & Anfinsen, P. A. Chemical dynamics in proteins: the photoisomerization of retinal in bacteriorhodopsin. *Science* **279**, 1886–1891 (1998).
- Humphrey, W., Lu, H., Logonov, I., Werner, H.-J. & Schulten, K. Three electronic state model of the primary phototransformation of bacteriorhodopsin. *Biophys. J.* **75**, 1689–1699 (1998).
- Zhong, Q., Ruhman, S. & Ottolenghi, M. Reexamining the primary light-induced events in bacteriorhodopsin using a synthetic C13=C14-locked chromophore. *J. Am. Chem. Soc.* **118**, 12828–12829 (1996).
- Ye, T. *et al.* On the nature of the primary light-induced events in bacteriorhodopsin: ultrafast spectroscopy of native and C13=C14 locked pigments. *J. Phys. Chem. B* **103**, 5122–5130 (1999).
- Song, L. & El-Sayed, M. A. Primary step in bacteriorhodopsin photosynthesis: bond stretch rather than angle twist of its retinal excited-state structure. *J. Am. Chem. Soc.* **120**, 8889–8890 (1998).
- Vreven, T. *et al.* Ab initio photoisomerization dynamics of a simple retinal chromophore model. *J. Am. Chem. Soc.* **119**, 12687–12688 (1997).
- Garavelli, M., Negri, F. & Olivucci, M. Initial excited-state relaxation of the isolated 11-*cis* protonated Schiff base of retinal: Evidence for in-plane motion from ab initio quantum chemical simulation of the resonance Raman spectrum. *J. Am. Chem. Soc.* **121**, 1023–1029 (1999).
- Du, M. & Fleming, G. R. Femtosecond time resolved fluorescence spectroscopy of bacteriorhodopsin: direct observation of excited state dynamics in the primary step of the proton pump cycle. *Biophys. Chem.* **48**, 101–111 (1993).

18. Peteanu, L. A., Shoenlein, R. W., Wang, Q., Mathies, R. A. & Shank, C. V. The first step in vision occurs in femtoseconds: complete blue and red spectral studies. *Proc. Natl Acad. Sci. USA* **90**, 11762–11766 (1993).
19. Shoenlein, R. W., Peteanu, L. A., Wang, Q., Mathies, R. A. & Shank, C. V. Femtosecond dynamics of cis-trans isomerization in a visual pigment analog: isorhodopsin. *J. Phys. Chem.* **97**, 12087–12092 (1993).
20. Mathies, R. A., Brito Cruz, C. H., Pollard, W. T. & Shank, C. V. Direct observation of the femtosecond excited-state cis-trans isomerization in bacteriorhodopsin. *Science* **240**, 777–779 (1988).
21. Atkinson, G. H., Brack, T. L., Blanchard, D. & Rumbles, G. Picosecond time-resolved resonance Raman spectroscopy of the initial trans to cis isomerization in the bacteriorhodopsin photocycle. *Chem. Phys.* **131**, 1–15 (1989).
22. van den Berg, R., Jang, D. J., Bittling, H. C. & El-Sayed, M. A. Subpicosecond resonance Raman spectra of the early intermediates in the photocycle of bacteriorhodopsin. *Biophys. J.* **58**, 135–141 (1990).
23. Doig, S. J., Reid, P. J. & Mathies, R. A. Picosecond time-resolved resonance Raman spectroscopy of bacteriorhodopsin J, K, and KL intermediates. *J. Phys. Chem.* **95**, 6372–6379 (1991).
24. Diller, R. *et al.* Femtosecond time-resolved infrared laser study of the J-K transition of bacteriorhodopsin. *Chem. Phys. Lett.* **241**, 109–115 (1995).
25. Pollard, W. T. *et al.* Theory of dynamic absorption spectroscopy of nonstationary states. 4. Application to 12-fs resonant impulsive Raman spectroscopy of bacteriorhodopsin. *J. Phys. Chem.* **96**, 6147–6158 (1992).
26. Bardeen, C. J., Wang, Q. & Shank, C. V. Femtosecond chirped pulse excitation of vibrational wave packets in LD690 and bacteriorhodopsin. *J. Phys. Chem. A* **102**, 2759–2766 (1998).
27. Eyring, G., Curry, B., Broek, A., Lugtenburg, J. & Mathies, R. Assignment and interpretation of hydrogen out-of-plane vibrations in the resonance Raman spectra of rhodopsin and bathorhodopsin. *Biochemistry* **21**, 384–393 (1982).
28. Myers, A. B., Harris, R. A. & Mathies, R. A. Resonance Raman excitation profiles of bacteriorhodopsin. *J. Chem. Phys.* **79**, 603–613 (1983).
29. Kobayashi, T., Shirakawa, A., Matsuzawa, H. & Nakanishi, H. Real-time vibrational mode-coupling associated with ultrafast geometrical relaxation in polydiacetylene induced by sub-5-fs pulses. *Chem. Phys. Lett.* **321**, 385–393 (2000).
30. Kobayashi, T. & Shirakawa, A. Tunable visible and near-infrared pulse generator in a 5 fs regime. *Appl. Phys. B* **70**, S239–S246 (2000).

Acknowledgements

We thank J. Watson and M. Murao for careful reading of the manuscript. This work was partially supported by the Research for the Future program run by the Japan Society for Promotion of Science (T.K.), the Special Coordination Funds ("Molecular Sensors for Aero-Thermodynamic Research"; H.O.) and Scientific Research (H.O.) of the Ministry of Education, Culture, Sports, Science and Technology.

Correspondence and requests for materials should be addressed to T.K. (e-mail: kobayashi@phys.s.u-tokyo.ac.jp).

Systematic distortions in world fisheries catch trends

Reg Watson* & Daniel Pauly*

* Fisheries Centre, 2204 Main Mall, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Over 75% of the world marine fisheries catch (over 80 million tonnes per year) is sold on international markets, in contrast to other food commodities (such as rice)^{1,2}. At present, only one institution, the Food and Agriculture Organization of the United Nations (FAO) maintains global fisheries statistics. As an inter-governmental organization, however, FAO must generally rely on the statistics provided by member countries, even if it is doubtful that these correspond to reality. Here we show that misreporting by countries with large fisheries, combined with the large and widely fluctuating catch of species such as the Peruvian anchoveta, can cause globally spurious trends. Such trends influence unwise investment decisions by firms in the fishing sector and by banks, and prevent the effective global management of international fisheries.

World fisheries catches have greatly increased since 1950, when the FAO of the United Nations began reporting global figures³. The reported catch increases were greatest in the 1960s, when the traditional fishing grounds of the North Atlantic and North Pacific became fully exploited, and new fisheries opened at lower latitudes

and in the Southern Hemisphere. Global catches increased more slowly after the 1972 collapse of the Peruvian anchoveta fishery⁴, the first fishery collapse that had repercussions on global supply and prices of fishmeal (Fig. 1a). Even taking into account the variability of the anchoveta, global catches were therefore widely expected to plateau in the 1990s at values of around 80 million tonnes, especially as this figure, combined with estimated discards of 16–40 million tonnes⁵, matched the global potential estimates published since the 1960s (ref. 6). Yet the global catches reported by the FAO generally increased through the 1990s, driven largely by catch reports from China.

These reports appear suspicious for the following three reasons: (1) The major fish populations along the Chinese coast for which assessments were available had been classified as overexploited decades ago, and fishing effort has since continued to climb^{7,8}; (2) Estimates of catch per unit of effort based on official catch and effort statistics were constant in the Yellow, east China and south China seas from 1980 to 1995 (ref. 9), that is, during a period of continually increasing fishing effort and reported catches, and in contrast to declining abundance estimates based on survey data⁷; (3) Re-expressing the officially reported catches from Chinese waters on

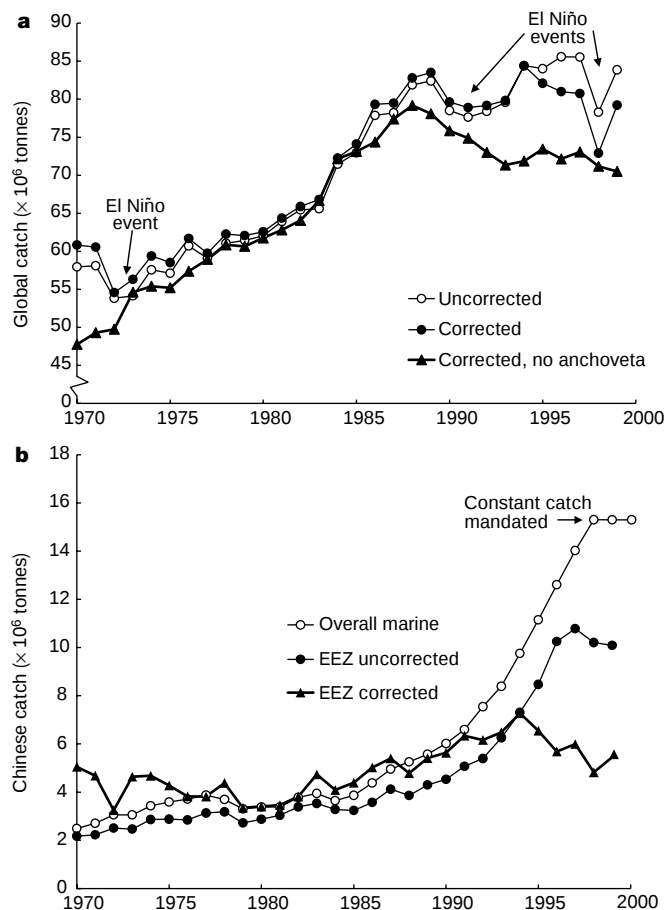


Figure 1 Time series of global and Chinese marine fisheries catches (1950 to present). **a**, Global reported catch, with and without the highly variable Peruvian anchoveta. Uncorrected figures are from FAO (ref. 3); corrected values were obtained by replacing FAO figures by estimates from **b**. The response to the 1982–83 El Niño/Southern Oscillation (ENSO) is not visible as anchoveta biomass levels, and hence catches were still very low from the effect of the previous ENSO in 1972 (ref. 4). **b**, Reported Chinese catches (from China's exclusive economic zone (EEZ) and distant water fisheries) increased exponentially from the mid-1980s to 1998, when the 'zero-growth policy' was introduced. The corrected values for the Chinese EEZ were estimated from the general linear model described in the Methods section.

a per-area basis leads to catches far higher than would be expected by comparison with similar areas (in terms of latitude, depth, primary production) in other parts of the world.

We investigated the third reason in some detail by generating world fisheries catch maps on the basis of FAO fisheries catch statistics for every year since 1950 (see Fig. 2a for a 1998 example). A statistical model was used to describe relationships between oceanographic and other factors, and the mapped catch. Most high-catch areas of the world were correctly predicted by the model. These areas typically had very high primary productivity rates driven by coastal upwellings, like those off Peru, supporting a large reduction fishery for the planktivorous anchoveta *Engraulis ringens*⁴. The exception was the waters along the Chinese coast. Here, the high catches could not be explained by the model using oceanographic or other factors. Yet the catch statistics provided to FAO by China have continued to increase from the mid-1980s until 1998 when, under domestic and international criticism, the government proclaimed a 'zero-growth policy' explicitly stating that reported catches would remain frozen at their 1998 value (Fig. 1b)¹⁰.

Mapping the difference between expected (that is, modelled) catches and those mapped from reported statistics showed large areas along the Chinese coast that had differences greater than 5 tonnes km⁻² year⁻¹. Overall, the statistical model for 1999 predicted a catch of 5.5 million tonnes, against an official report of 10.1 million tonnes (see Fig. 1b for earlier years). Although China was not the only FAO member country reporting relatively high catches, their large absolute value strongly affects the global total.

For a number of obvious reasons, fishers usually tend to under-report their catches, and consequently, most countries can be presumed to under-report their catches to FAO. Thus we wondered why China should differ from most other countries in this way. We believe that explanation lies in China's socialist economy, in which the state entities that monitor the economy are also given the task of increasing its output¹¹. Until recently, Chinese officials, at all levels, have tended to be promoted on the basis of production increases from their areas or production units¹¹. This practice, which originated with the founding of the People's Republic of China in 1949, became more widespread since the onset of agricultural reforms

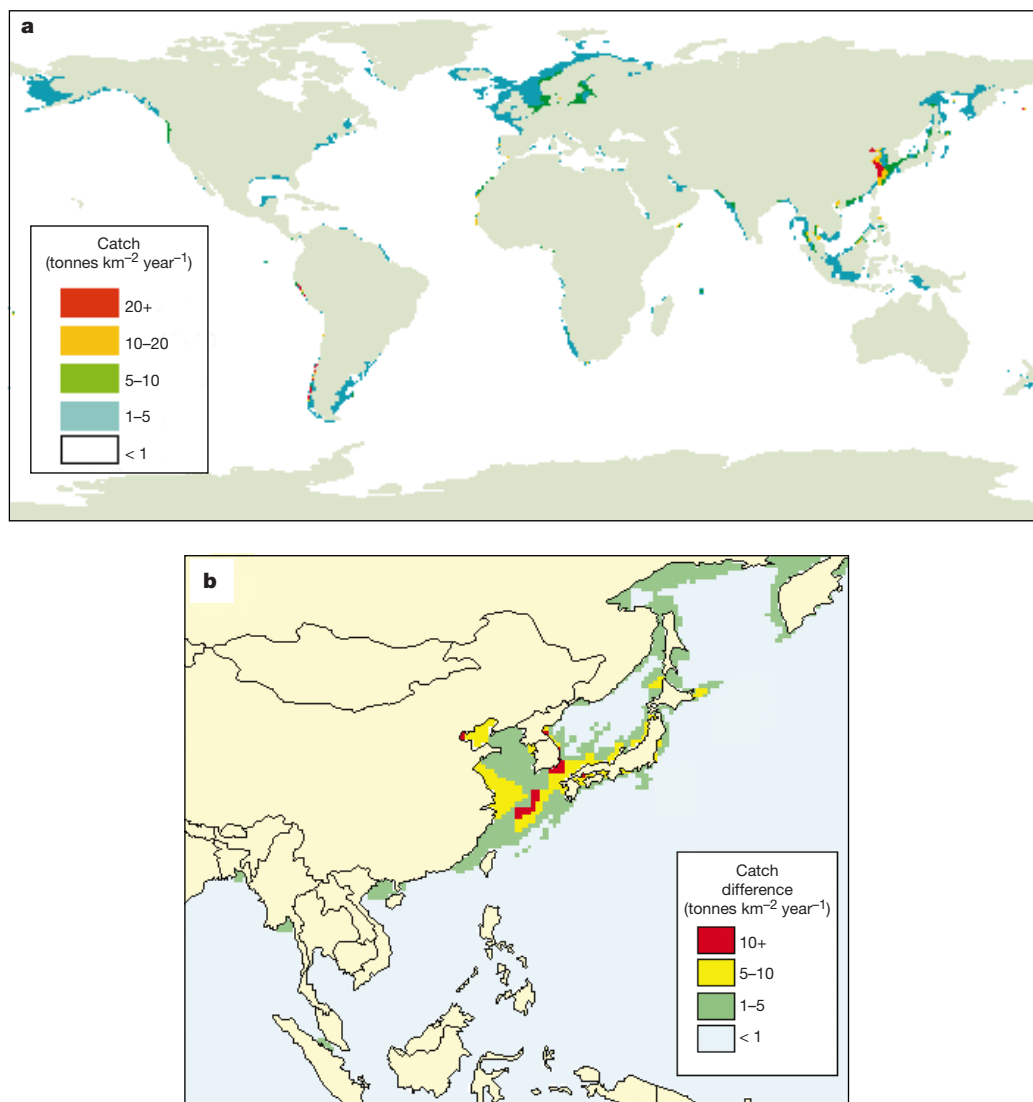


Figure 2 Maps used to correct Chinese marine fisheries catch in Fig. 1b. **a**, Map of global catches reported by FAO for 1998, generated by the rule-based algorithm described in the Methods section. We note the anomalously high values along the Chinese coast, comparable in intensity (not area covered) to the extremely productive

Peruvian upwelling system. **b**, Map of differences in southeast and northeast Asia between the catches reported in **a** and those predicted by the model described in the Methods section.

that freed the agricultural sector from state directives in the late 1970s (refs 10, 11).

The Chinese central government appears to be well aware of this problem, and its 1998 'zero-growth policy' was partly intended to prevent over-reporting. Thus, the official fisheries catches for 1999–2000 are precisely the same as in 1998 (Fig. 1b), and will be for the next few years. Such measures, although well motivated, do not inspire confidence in official statistics, past or present.

The substitution of the more realistic estimated series of Chinese catches into the FAO fisheries statistics led to global catch estimates which, although fluctuating, have tended to decline by 0.36 million tonnes year⁻¹ since 1988 (rather than increase by 0.33 million tonnes year⁻¹, as suggested by the uncorrected data). The global downward trend becomes clearer when the catches of a single species, the Peruvian anchoveta, which is known to be affected by El Niño/Southern Oscillation events, is subtracted (see Fig. 1a). In this case, a significant ($P < 0.01$), and so far undocumented downward trend of 0.66 million tonnes year⁻¹ becomes apparent for all other species and fisheries. This is consistent with other accounts of worldwide declines of fisheries^{12,13}.

Ironically, it is likely that, at the lowest levels (individual fishers), catches are under-reported in China as elsewhere in the world. The production targets caused these reports to be exaggerated. At some times these two distortions may perhaps have cancelled each other out, and an accurate report of catches may have been submitted to FAO. Since the early 1990s, however, the exaggerations have apparently far exceeded any initial under-reporting.

The greatest impact of inflated global catch statistics is the complacency that it engenders. There seems little need for public concern, or intervention by international agencies, if the world's fisheries are keeping pace with people's needs. If, however, as the adjusted figures demonstrate, the catches of world fisheries are in general decline, then there is a clear need to act. The oceans should continue to provide for a substantial portion of the world's protein needs. The present trends of overfishing, wide-scale disruption of coastal habitats and the rapid expansion of non-sustainable aquaculture enterprises¹⁴, however, threaten the world's food security. □

Methods

Data processing involved a disaggregation of global fisheries catch statistics firstly into detailed taxonomic groups, and then into fine-scale spatial cells (a half-degree of latitude by a half-degree of longitude), using a variety of databases and systematic rules¹⁵. The spatially disaggregated catches provided the basis for a general linear model of fisheries catches (see below). The model predicted the likely catches in the spatial cells in the Chinese exclusive economic zone (EEZ), thus providing an estimate of Chinese catches (including Hong Kong and Macau, but excluding Taiwan).

Data sources

Fisheries catch statistics were provided by the FAO (FishStat³ and 'Atlas of Tuna and Billfish Catches', <http://www.fao.org/fi/atlas/tunabill/english/home/htm>). The spatial cells were described by depth (US National Geophysical Data Center), primary productivity (Joint Research Centre of the European Commission Space Applications Institute—Marine Environment Unit, http://www.gmes.jrc.it/download/kyoto_prot/glob.marine.pdf), biogeochemical provinces¹⁶, the presence of ice (US National Snow and Ice Data Center, <http://www.nsidc.org>), surface temperature (NOAA's Marine Atlas, http://www.nodc.noaa.gov/OC5/data_woa.html), and an upwelling index calculated for each cell by multiplying negative deviations in surface temperature (from the average for that latitude and ocean) by the primary productivity in that cell. Fishing access rights were determined using maps of the exclusive economic zones (EEZ) of coastal states¹⁷ and a database of fishing access agreements¹⁸.

Taxonomic disaggregation

The fisheries statistics of several nations commonly include a large fraction of catches in 'miscellaneous' categories. Chinese catches so reported were disaggregated on the basis of the breakdown provided by its two nearest maritime neighbours with detailed marine fisheries statistics (Taiwan and South Korea)¹⁵. Assigning catches to lower taxa allowed the use of biological information in the spatial disaggregation process.

Spatial disaggregation

A database of the global distribution of commercial fisheries species was developed using information from a variety of sources including the FAO, FishBase¹⁹ and experts on various

resource species or groups. Some distributions were specific; others provided depth or latitudinal limits, or simple presence/absence data. The spatial disaggregation process determined the intersection set of spatial cells within the broad statistical area for which the statistics were provided to FAO, the global distribution of the reported species, and the cells to which the reporting nation had access through fishing agreements¹⁵. The reported catch tonnage was then proportioned within this set of cells.

Catch predictions

A general linear model was developed in the software package S-Plus²⁰. The model relates log fisheries catch (in tonnes km⁻² year⁻¹) for each cell (the dependent variable) to depth, primary productivity, ice cover, surface temperature, latitude, distance from shore, upwelling index (the continuous predictor variables), 33 oceanic biogeochemical provinces and one global coastal 'biome'¹⁶ including most of the area covered by the world's EEZs, including China's (the categorical predictor variables). Fishing effort was not used in the prediction and catches were assumed to be generally close to their maximum biologically sustainable limits. The additive and variance stabilizing transformation (AVAS) routine of S-Plus²⁰ was used to identify transformations ensuring linearity between the dependent and explanatory variables, and the model was then used to predict the catch from each spatial cell. Those from Chinese waters were combined, then compared with the catches obtained from the rule-based spatial disaggregation described above.

Trend analyses

The estimates of recent trends of global catch were estimated by linear regression of catch versus year, for the period from 1988 (highest catches, anchoveta excluded) to 1999 (last year with FAO data), for uncorrected global marine catches, global marine catches adjusted for Chinese over-reporting, and adjusted catches minus the catch of Peruvian anchoveta.

Received 30 July; accepted 28 September 2001.

- Food and Agriculture Organization. *Fisheries Trade Flow (1995–1997)* 1–330 (Fisheries Information, Data and Statistics Unit, FAO Fisheries Circular No. 961, Rome, 2000).
- IRRI. in *Rice Almanac* (ed. Maclean, J.) 2nd edn, 31–37 (International Rice Research Institute, Los Baños, Philippines, 1997).
- Food and Agriculture Organization. FISHSTAT Plus. Universal software for fishery statistical time series. Version 2.3. (Fisheries Department, Fisheries Information, Data and Statistics Unit, Rome, 2000).
- Muck, P. in *The Peruvian Upwelling Ecosystem: Dynamics and Interactions* (eds Pauly D., Muck, P., Mendo, J. & Tsukayama, L.) 386–403 (International Centre for Living Aquatic Resource Management, Makati, Philippines, 1989).
- Alverson, D. L., Freeberg, M., Pope, J. & Murawski, S. A Global Assessment of Fisheries By-catch and Discards: A Summary Overview. 1–233 (FAO Fisheries Technical Paper 339, Rome, 1994).
- Pauly, D. One hundred million tonnes of fish, and fisheries research. *Fish. Res.* **25**, 25–38 (1996).
- Tang, Q. in *Biomass Yields and Geography of Large Marine Ecosystems* (eds Sherman, K. & Alexander, L. M.) 7–35 (AAAS Selected Symp. 111, Westview, Boulder, 1989).
- Huang, B. & Walters, C. J. Cohort analysis and population dynamics of large yellow croaker in the China Sea. *N. Am. J. Fish. Man.* **3**, 295–305 (1983).
- Chen, W. Marine Resources: Their Status of Exploitation and Management in the People's Republic of China. 1–60 (FAO Fisheries Circular No. 950, Rome, 1999).
- Pang, L. & Pauly, D. in *The Marine Fisheries of China: Development and Reported Catches* (authors Watson, R., Pang, L. & Pauly, D.) 1–27 (Fisheries Centre Research Report 9(2), Univ. British Columbia, Vancouver, 2001); also at <http://fisheries.ubc.ca/Reports/china.pdf>.
- Kwong, L. *The Political Economy of Corruption in China* 1–75 (M. E. Sharpe, Armonk, New York, 1997).
- Botsford, L. W., Castilla, J. C. & Peterson, C. H. The management of fisheries and marine ecosystems. *Science* **277**, 509–515 (1997).
- Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. & Torres, F. C. Jr Fishing down marine food webs. *Science* **279**, 860–863 (1998).
- Naylor, R. L. et al. Effect of aquaculture on world fish supplies. *Nature* **405**, 1017–1024 (2000).
- Watson, R., Gelchu, A. & Pauly, D. in *Fisheries Impacts on North Atlantic Ecosystems: Catch, Effort, and National/Regional Data Sets* (eds Zeller, D., Watson, R. & Pauly, D.) (Fisheries Centre Research Report, Univ. British Columbia, Vancouver, in the press).
- Longhurst, A. *Ecological Geography of the Sea* 77–85 (Academic, San Diego, 1998).
- Veridian Information Solutions 2000. Global Maritime Boundaries Database CD. (Veridian, Fairfax, Virginia, 2000).
- Food and Agriculture Organization. Fisheries Agreements Register (FARISIS). 1–4 (Committee on Fisheries, 23rd session, Rome, 1999; COFI/pp/Inf.9E, 1998).
- Froese, R. & Pauly, D. (eds) FishBase 2000. Concepts, design and data sources. ((International Centre for Living Aquatic Resource Management, Los Baños, Philippines, 2000); 4CD-ROMs; updates on <http://www.fishbase.org>).
- S-Plus Guide to Statistics. Vol. 1 (Data Analysis Product Division, MathSoft, Seattle, 2000).

Acknowledgements

We thank V. Christensen for the upwelling index, and A. Gelchu for the species distribution shape files used here. We also thank our colleagues in the 'Sea Around Us' Project. This work was supported by the Pew Charitable Trusts through the 'Sea Around Us' Project, Fisheries Centre, University of British Columbia. D.P. also acknowledges support from the National Science and Engineering Council of Canada.

Correspondence and requests for materials should be addressed to R.W. (e-mail: r.watson@fisheries.ubc.ca).

Demography of the endangered North Atlantic right whale

Masami Fujiwara & Hal Caswell

Biology Department, MS34, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA

Northern right whales (*Eubalaena glacialis*) were formerly abundant in the northwestern Atlantic, but by 1900 they had been hunted to near extinction. After the end of commercial whaling the population was thought to be recovering slowly; however, evidence indicates that it has been declining since about 1990 (ref. 1). There are now fewer than 300 individuals, and the species may already be functionally extinct^{2,3} owing to demographic stochasticity or the difficulty of females locating mates in the vast Atlantic Ocean (Allee effect⁴). Using a data set containing over 10,000 sightings of photographically identified individuals we estimated trends in right whale demographic parameters since 1980. Here we construct, using these estimates, matrix population models allowing us to analyse the causes of right whale imperilment. Mortality has increased, especially among mother whales, causing declines in population growth rate, life expectancy and the mean lifetime number of reproductive events between the period 1980–1995. Increased mortality of mother whales can explain the declining population size, suggesting that the population is not doomed to extinction as a result of the Allee effect. An analysis of extinction time shows that demographic stochasticity has only a small effect, but preventing the deaths of only two female right whales per year would increase the population growth rate to replacement level.

Conservation biology uses population models to assess population performance, to diagnose the causes of poor performance, to prescribe management interventions, and to make prognoses of population viability (see chapter 18 of ref. 5). We have developed a stage-structured matrix population model^{5–7} that addresses all four of these tasks. This matrix model uses the life cycle shown in Fig. 1 (a similar model has been successfully applied to a killer whale

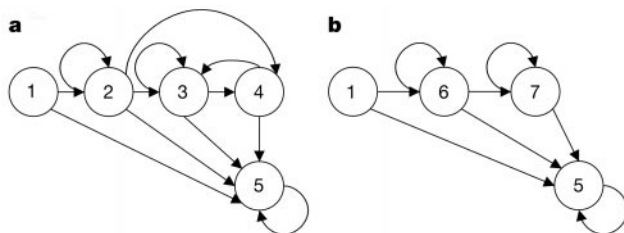


Figure 1 Life cycle graphs of female (a) and male (b) right whales. Numbers represent different stages: 1, calf; 2, immature female; 3, mature female; 4, mature females with newborn calves (mothers); 5, dead; 6, immature male; 7, mature male. Each arrow represents a possible transition in stage from one year to the next; the arrows going to stage 5 represent stage-specific mortalities. A calf is an individual that was sighted along with its mother. An immature is an individual known to be less than 9 years old. Mature individuals are known to be at least 9 years old, or in the case of females, have been sighted previously with a calf. Mothers are females that are sighted with a newborn offspring. If the sex of an individual is unknown, we assume it has an equal chance of being either female or male, as our data contain almost equal numbers of individuals (141 and 143) known to be female and male, respectively. Maturity status (whether an individual is immature or mature) is unknown in 22% of sightings. When maturity status was unknown, we estimated the probability that the individual was immature (0.30 for males, 0.87 for females) using the method described in ref. 10. These probabilities were used in stage-assignment matrices¹⁰ for likelihood calculations.

population⁸). The parameters in the matrix model were defined by a series of statistical models, the parameters of which were estimated by applying stage-structured mark–recapture analysis^{5,9,10} to photographic identification data collected by the New England Aquarium (NEA)¹¹. The best model was selected using Akaike Information Criteria (AIC)^{12–14} (see Methods).

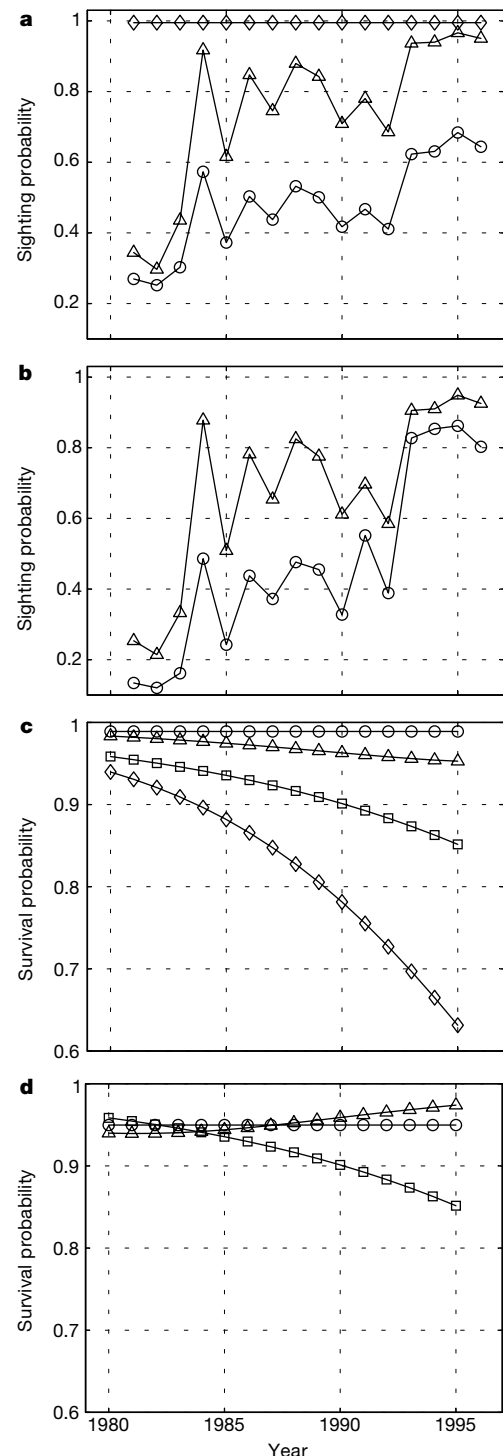


Figure 2 Stage-specific sighting and survival probabilities for males and females. a, b, Stage-specific sighting probabilities of the best model (M_2) for females (a) and males (b). c, d, Stage-specific survival probabilities (M_2) for females (c) and males (d). Squares, calf; triangles, immature; circles, mature; diamonds, mothers.

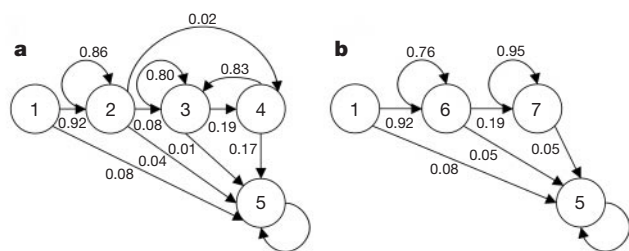


Figure 3 Constant transition probabilities estimated using the best sighting model (M_1). Life cycle graphs of female (a) and male (b) right whales. The numbers in the circles are the same as in Fig. 1.

In the resulting model, the sighting probability of most stages depends on both northern and southern effort levels (Fig. 2a, b; see Methods for definition). This is consistent with a previous study¹, which found a significant correlation between total effort and sighting probability in a simpler model that did not distinguish sex or stages. In contrast to other stages, mothers have a consistently high sighting probability (Fig. 2a). This suggests that they are easier to sight than other stages, and that the survey effort in the two regions has been sufficient to detect almost all births with high probability.

Using the best sighting probability model, we estimated two transition models. One model assumed time-invariant transition probabilities (M_1) and the other was the best time-varying model according to the AIC criterion (M_2). The time-invariant model (M_1 , $b_{ji} = 0$ for all j and i in equation (2)) gives a weighted, time-averaged picture of transition probabilities. In this model, survival probability is low for calves, higher for immature and mature females and much lower for mothers (Fig. 3). According to the best time-varying model (M_2), the transition probabilities of mature females and males have been constant, whereas the other stages have transition probabilities that change with time. The survival probability of mothers shows the most pronounced decline over time (Fig. 2c, d). This trend is statistically significant, as the slope parameter (b_{34}) in the polychotomous logistic function (equation (2)) is significantly below 0 (Fig. 4).

A series of population projection matrices, A_t , was constructed by augmenting the female transition probability matrix with elements describing reproduction¹⁰ (see Methods). These matrices were used to calculate population growth rate, life expectancy and expected lifetime number of reproductive events experienced by a female.

The asymptotic population growth rate (λ) is given by the dominant eigenvalue of the population projection matrix. When $\lambda > 1$, the population is asymptotically growing; when $\lambda < 1$, the population is asymptotically declining. Therefore, λ is an important indicator of the status of a population. If we assume all demographic parameters have been constant from 1980 to 1995 (M_1), we find that $\lambda = 1.01$ (95% confidence interval = [1.00, 1.02]; the confidence interval was estimated using the estimated covariance of parameters by taking the inverse of the Hessian matrix¹⁰), suggesting a population increase of 1% per year, on average, between 1980 and 1995. However, time-specific asymptotic growth rates (λ_t)—calculated from the time-varying matrices A_t using model M_2 —declined from $\lambda_{1980} = 1.03$ to $\lambda_{1995} = 0.98$ (Fig. 5c). Population growth rate declined below 1 at around 1992. If the 1995 population growth rate were maintained, the population would go extinct.

The decline in λ_t is the net result of all the changes in the vital rates (a collective term for transition, survival, fertility and mortality rates). We determined how much the decline in each vital rate contributed to the decline in λ_t , using a life table response experiment (LTRE) analysis^{5,6}. This analysis decomposes the changes in λ_t into contributions from each entry in A_t . Figure 5d shows the sum of the contributions of all matrix entries involving each stage. The

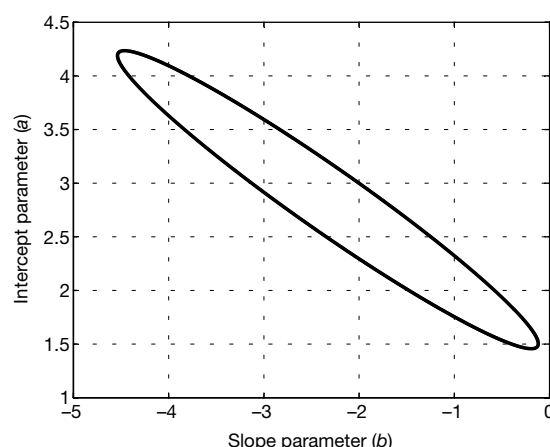


Figure 4 Joint 95% profile likelihood confidence region for the logistic parameters of survival probability of mothers.

results of this analysis show that the decline in λ_t is caused mainly by the decline in the survival probability of mothers (Fig. 5d).

The life expectancy at birth is the mean age at which individuals born in a given year would die if conditions of that year were maintained. It can be calculated by treating the transition matrix as an absorbing Markov chain and calculating the mean time to absorption (that is, death)⁵. During the early 1980s, the life expectancy of females was twice that of males (Fig. 5a). This may be typical of large cetaceans. (Killer whales (*Orcinus orca*), for example, exhibit a similar difference in life expectancy between females and males¹⁵.) Female life expectancy has declined from about 51.8 years in 1980 to about 14.5 years in 1995. Until recently, life expectancy of females has exceeded that of males, but that is no longer true owing to the reduced survival probability of females.

The expected number of reproductive events during a female's lifetime⁵ has declined from about 5.27 in 1980 to about 1.26 in 1995 (Fig. 5b). A mature female could once expect to reproduce 6.48 times; that number is now about 1.80.

The growth rate projections in Fig. 5c are deterministic. The right whale population is small, however, and the population of any single stage is even smaller. Because of this, it has been suggested that population projections should include demographic stochasticity (chance fluctuations due to the random fates of a small number of individuals; see ref. 5). To evaluate the effect of demographic stochasticity, we transformed the matrix A_{1995} into a multitype branching process⁵, and calculated the probability distribution of times to extinction (Fig. 6). We compared this distribution with the deterministic prediction obtained by using A_{1995} and defining extinction as the time when total population size reached 1 individual. For both calculations, we used an initial condition of 150 females distributed according to the stable stage distribution. This simulates the hypothetical situation in which the vital rates of 1995 remain constant and neither environmental trends nor environmental fluctuations affect them.

The mean time to extinction under demographic stochasticity is 208 years (similar to the estimate of 191 years in ref. 1). The deterministic time to extinction is 245 years. Thus demographic stochasticity reduces expected extinction time by 15%. Figure 6 gives the complete distribution of extinction times; there is a 5% chance of extinction within 130 years and a 25% chance of extinction within 165 years. These calculations exclude other factors such as continued declining survival trends, environmental stochasticity and Allee effects, all of which would hasten extinction. Thus the expected time of 208 years should be considered an upper bound.

Right whale conservation efforts are directed towards reducing mortality due to entanglement and ship collisions. Because the

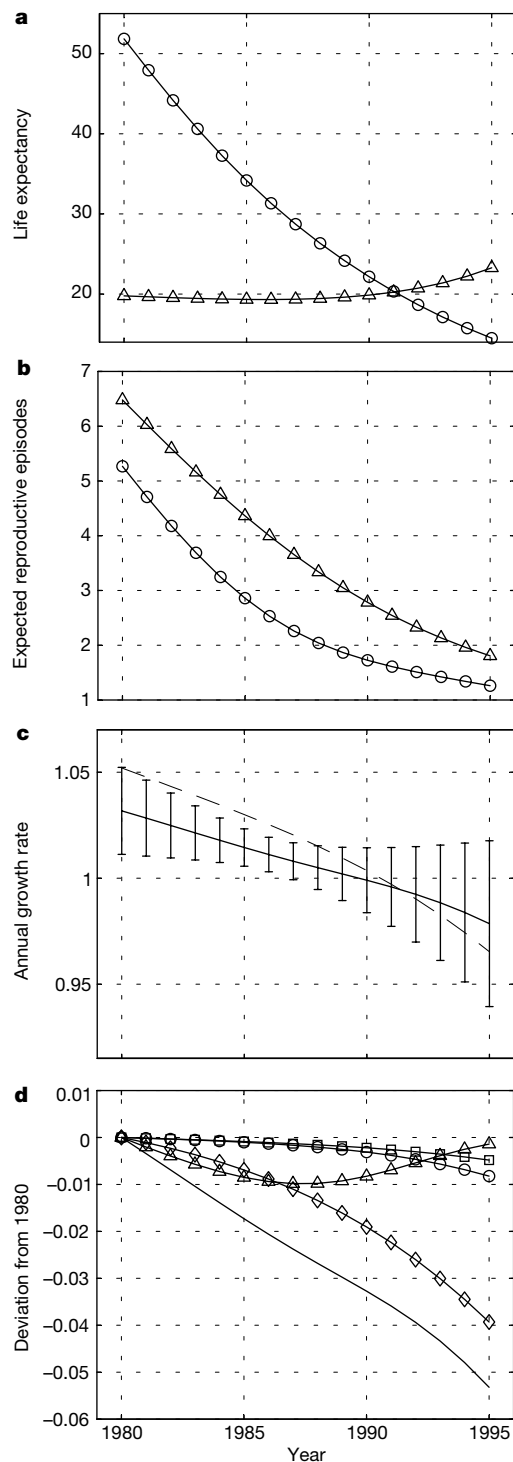


Figure 5 Demographic parameters calculated from the time-varying matrices A_t . **a**, The mean life expectancy at birth for males (triangles) and females (circles). **b**, The mean number of lifetime reproductive episodes, estimated from birth (circles) and from maturity (triangles). **c**, Asymptotic population growth rates λ_t (solid line) calculated from the population projection matrices A_t produced by the best model (M_2) and from the unstructured model of ref. 1 (dotted line). The error bars were standard errors, which were calculated using the series approximation to the variance of λ and the covariance matrix obtained from the information matrix. **d**, The result of a LTRE decomposition analysis for the trend in λ . The demographic rates in 1980 are used as the reference matrix, and the contributions of all matrix entries involving each stage were summed. The solid line shows the actual trend in λ , measured as a deviation from the value in 1980. It is closely approximated by the sum of all the other lines. Squares, calf; triangles, immature; circles, mature; diamonds, mothers.

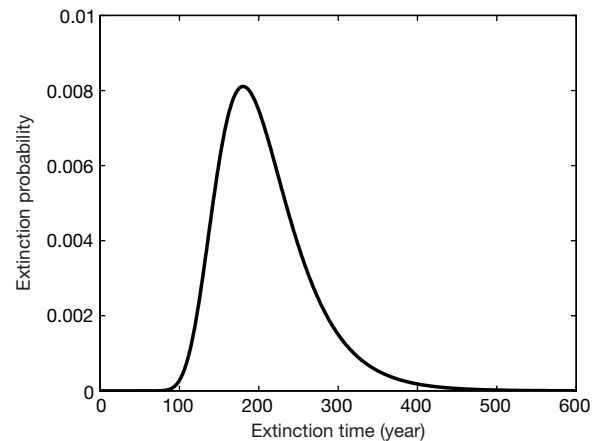


Figure 6 The probability distribution of time to extinction assuming demographic stochasticity. The distribution was calculated from a multi-type branching process, treating transitions and reproduction as independent events (multinomial and Bernoulli, respectively; see ref. 5).

population is so small, a single death represents a significant mortality rate. Conversely, significant reduction in mortality rate can be obtained by preventing just a few deaths. Figure 7 shows the effect on λ of preventing female deaths, using the vital rates of 1995 and assuming a starting population of 150 females distributed according to the stable stage distribution. The results indicate that prevention of just two female deaths every year would suffice to increase λ to 1. This gives a sense of the magnitude of the initial management action needed for the protection of the population. However, as population size increases owing to successful management, more deaths will have to be prevented to maintain a positive population growth rate because the number of deaths that translates into a given mortality rate also increases with the population size.

The causes behind the decline in survival probability of mothers are still unknown; however, collisions with ships, entanglement with fishing gear and changes in food availability due to climate fluctuations are suspected to contribute towards mortality^{16,17}. Although right whales are distributed from the coast of northern Florida to the Bay of Fundy, it is primarily females and calves that are sighted in the calving ground off the coast of Florida and Georgia¹⁸. The calving ground is close to shipping lanes where large vessel traffic has increased since 1980 (ref. 19). More than 60% of North Atlantic right whales have scars from entanglement in fishing gear such as lobster pots and sink gillnets¹⁷. We have found a significant rank correlation between crude survival probability¹ (the survival probability of individuals without distinguishing stage differences) and North Atlantic Oscillation index (NAO), which is often used as an index for climate fluctuations²⁰. Studies that focus on the effects of these three factors are urgently needed.

This study combines multi-stage mark-recapture methods and matrix population model analyses. Both methods incorporate differences in vital rates that are experienced by different stages within a population. Multi-stage mark-recapture analysis also incorporates stage-specific heterogeneity in sighting probability within the population. We succeeded in identifying a life cycle stage that is experiencing reduced survival probability, and were able to use the model to document the implications of that reduced survival probability for population growth.

Like any mark-recapture analysis, ours can be influenced by heterogeneity in sighting probability among individuals within a stage. If the sighting probability of mature females (stage 3) was heterogeneous, the survival probability of mothers towards the end of the study period would be underestimated. However, heterogeneity in sampling alone cannot explain the observed decline in the

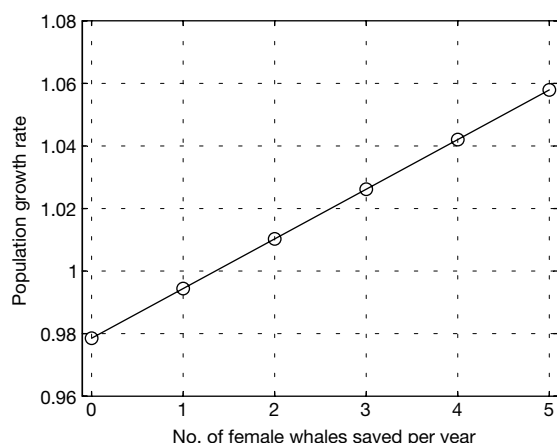


Figure 7 The predicted population growth rate that would result from preventing deaths of females regardless of their stage.

survival probability. Furthermore, during the winter of 1996, an unusually large number of confirmed mortalities were reported in the southeast United States, which is the only known calving ground¹⁷. This evidence suggests that the declining trend in the survival of mothers is both real and a great concern.

The North Atlantic right whale is currently seriously endangered owing to declining survival probability, especially among mothers. This finding, together with the unprecedented calf production in the spring of 2001 (ref. 21), suggests that Allee effects are not yet limiting this population. Our analysis also shows that the population was experiencing a positive growth rate in the early 1980s (Fig. 5c). This implies that it is not necessary to return the vital rates to those of the pre-whaling period to obtain a positive growth rate. There is every reason to hope that prompt management intervention can improve survival enough to permit the recovery of the North Atlantic right whale. □

Methods

Photographic identification data

Since 1980, the NEA has accumulated photographs of approximately 350 right whales taken on more than 10,000 sighting occasions. Individual right whales can be recognized by natural markings such as scars and callosity patterns. On the basis of these identifications the NEA has constructed a database of individual sighting histories, which consist of information on whether or not each individual was sighted at least once in each year. These data can be treated as if individuals were marked on the occasion of their first identifications, and recaptured at subsequent sightings. We used the data from 1980 to 1996, which were complete at the beginning of this analysis.

Statistical models

Sighting probabilities of each stage and transition probabilities among stages were estimated by fitting a series of statistical models to the sighting history data by maximum likelihood. The statistical models describe sighting probability as a function of sampling effort in two regions. The northern effort level is measured by the total annual survey days in Massachusetts Bay, Great South Channel, Bay of Fundy, and in Brown's Bank; the southern effort level is measured by the total annual survey days off the coast of Florida and Georgia. The dependence of sighting probability on the northern effort, the southern effort, both efforts, or neither effort was modelled with binary logistic functions. Let $s_i(t)$ be the sighting probability of stage i at time t . Then

$$s_i(t) = \frac{\exp(x_i + y_i e_{1,t} + z_i e_{2,t})}{1 + \exp(x_i + y_i e_{1,t} + z_i e_{2,t})} \quad (1)$$

where x_i is an intercept parameter, and y_i and z_i are slope parameters associated with northern ($e_{1,t}$) and southern ($e_{2,t}$) effort levels, respectively.

We modelled the transition probabilities of each stage as polychotomous logistic functions of time^{10,22}. Let $p_{ji}(t)$ be the transition probability from stage i at time t to j at time $t+1$. Then

$$p_{ji}(t) = \frac{\exp(a_{ji} + b_{ji}t)}{1 + \sum_j \exp(a_{ji} + b_{ji}t)} \quad (2)$$

where a_{ji} and b_{ji} are intercept and slope parameters, respectively. When all the slope parameters are set to zero, the transition probabilities are time-invariant.

Model selection

Model selection was done in the following sequence. First, 1,024 sighting models were created by including all possible combinations of effort levels for all possible combinations of stages (equation (1)). These models were combined with the transition model in equation (2) in which all probabilities are functions of time. We selected the best of these models using AIC¹². The AIC difference between the best and the second-best sighting models was about 2, indicating that the support for the best model relative to the second best model is high¹⁴. Therefore, we used only the best model. Because sighting probabilities of immature males and females did not differ significantly, by a likelihood ratio test, they were set equal. Using the resulting model for sighting, we selected the best time-varying transition matrix from all 64 models created by allowing the transition probabilities of all possible combination of stages to depend on time according to equation (2). The four best transition models had AIC differences from the best model of less than 2, indicating that the data provide some support for all these models. All of these models, however, had time-dependent survival probability of mothers.

Projection matrix

The projection matrix is

$$A_t = \begin{pmatrix} 0 & F_2 & F_3 & 0 \\ p_{21}(t) & p_{22}(t) & 0 & 0 \\ 0 & p_{32}(t) & p_{33}(t) & p_{34}(t) \\ 0 & p_{42}(t) & p_{43}(t) & 0 \end{pmatrix} \quad (3)$$

where

$$F_2 = 0.5p_{42}(t)p_{34}^{0.5}(t+1) \quad (4)$$

and

$$F_3 = 0.5p_{43}(t)p_{34}^{0.5}(t+1) \quad (5)$$

The first row of A_t describes reproduction; the other entries are transition probabilities. Consider F_2 . When a female moves from stage 2 to stage 4 (with probability $p_{42}(t)$), she gives birth; the newborn calf is female with probability 0.5. Although newborn calves have distinct markings, they are harder to distinguish individually than other stages. Therefore, calf survival is estimated from the point where the calf is seen sufficiently well to permit identification, which is not necessarily on its first sighting. To appear as a calf in stage 1 at $t+1$, the newborn calf must survive long enough to be catalogued. We assume calves are catalogued on average midway through their first year, and that the mother must survive this long (with probability $p_{34}^{0.5}(t+1)$) in order for the calf to survive. F_3 is similar.

Received 18 May; accepted 21 September 2001.

- Caswell, H., Fujiwara, M. & Brault, S. Declining survival probability threatens the North Atlantic right whale. *Proc. Natl Acad. Sci. USA* **96**, 3308–3313 (1999).
- Knowlton, A. R., Kraus, S. D. & Kenney, R. D. Reproduction in North Atlantic right whales (*Eubalaena glacialis*). *Can. J. Zool.* **72**, 1297–1305 (1994).
- Gerber, L. R., DeMaster, D. P. & Roberts, S. P. Measuring success in conservation. *Am. Sci.* **88**, 316–324 (2000).
- Allee, W. C. *Animal Aggregations* (Univ. Chicago Press, Chicago, 1931).
- Caswell, H. *Matrix Population Models: Construction, Analysis and Interpretation* 2nd edn (Sinauer Associates, Sunderland, Massachusetts, 2001).
- Caswell, H. *Matrix Population Models* (Sinauer Associates, Sunderland, Massachusetts, 1989).
- Tuljapourkar, S. & Caswell, H. *Structured-population Models in Marine, Terrestrial, and Freshwater Systems* (Chapman & Hall, New York, 1997).
- Brault, S. & Caswell, H. Pod-specific demography of killer whales (*Orcinus orca*). *Ecology* **74**, 1444–1454 (1993).
- Nichols, J. D., Sauer, J. R., Pollock, K. H. & Hestbeck, J. B. Estimating transition probabilities for stage-based population projection matrices using capture-recapture data. *Ecology* **73**, 306–312 (1992).
- Fujiwara, M. & Caswell, H. Estimating population projection matrices from multi-stage mark-recapture data. *Ecology* (in the press).
- Crone, M. J. & Kraus, S. D. *Right Whale (Eubalaena glacialis) in the Western North Atlantic: A Catalog of Identified Individuals* (New England Aquarium, Boston, Massachusetts, 1990).
- Akaike, H. in *International Symposium on Information Theory* 2nd edn (eds Petran, B. N. & Csaki, F.) 267–281 (Akademiai Kiado, Budapest, 1973).
- Lebreton, J.-D., Burnham, K. P., Clobert, J. & Anderson, D. R. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecol. Monogr.* **62**, 67–118 (1992).
- Burnham, K. P. & Anderson, D. R. *Model Selection and Inference: a Practical Information-theoretic Approach* (Springer, New York, 1998).
- Olesiuk, P. F., Bigg, M. A. & Ellis, G. M. in *Individual Recognition of Cetaceans: Use of Photo-identification and Other Techniques to Estimate Population Parameters* (eds Hammond, P. S., Mizroch, S. A. & Donovan, G. P.) 209–243 (Rep. Int. Whaling Comm. Special Issue 12, Cambridge, 1990).
- Kraus, S. D. Rates and potential causes of mortality in North Atlantic right whales (*Eubalaena glacialis*). *Mar. Mamm. Sci.* **6**, 278–291 (1990).
- Waring, G. T. et al. *U.S. Atlantic and Gulf of Mexico Marine Mammal Stock Assessments—1999*. NOAA Technical Memorandum NMFS-NE-153 (Northeast Fisheries Science Center, Woods Hole, Massachusetts, 1999).

18. Kraus, S. D., Prescott, J. H., Knowlton, A. R. & Stone, G. S. in *Right Whale: Past and Present Status* (eds Brownell, R. L. Jr, Best, P. B. & Prescott, J. H.) 145–151 (Rep. Int. Whaling Comm. Special Issue 10, Cambridge, 1986).
19. Knowlton, A. R. in *Shipping/Right Whale Workshop* (eds Knowlton, A. R., Kraus, S. D., Meck, D. F. & Mooney-Seus, M. L.) 31–36 (New England Aquarium Aquatic Forum Series Report 97–3, Massachusetts, 1997).
20. Planque, B. & Reid, P. C. Predicting *Calanus finmarchicus* abundance from a climatic signal. *J. Mar. Biol. Ass. UK* **78**, 1015–1018 (1998).
21. Holden, C. Whale baby boomlet. *Science* **291**, 429 (2001).
22. Hosmer, D. W. & Lemeshow, D. *Applied Logistic Regression* (Wiley, New York, 1989).

Acknowledgements

We thank S. Kraus, A. Knowlton, P. Hamilton and the NEA for data. We also thank S. Braut, M. Hill, P. Karciva, J.-D. Lebreton, M. Neubert, J. Nichols and the participants of the first Woods Hole Workshop on the Demography of Marine Mammals for discussions and suggestions. This project was funded by the Woods Hole Oceanographic Institution Sea Grant Program, the Rinehart Coastal Research Center, the David and Lucile Packard Foundation and The Robert W. Morse Chair.

Correspondence and requests for materials should be addressed to M.F. (e-mail: mfujiiwara@whoi.edu).

Transgenic DNA introgressed into traditional maize landraces in Oaxaca, Mexico

David Quist & Ignacio H. Chapela

Department of Environmental Science, Policy and Management, University of California, Berkeley, California 94720-3110, USA

Concerns have been raised about the potential effects of transgenic introductions on the genetic diversity of crop landraces and wild relatives in areas of crop origin and diversification, as this diversity is considered essential for global food security. Direct effects on non-target species^{1,2}, and the possibility of unintentionally transferring traits of ecological relevance onto landraces and wild relatives have also been sources of concern^{3,4}. The degree of genetic connectivity between industrial crops and their progenitors in landraces and wild relatives is a principal determinant of the evolutionary history of crops and agroecosystems throughout the world^{5,6}. Recent introductions of transgenic DNA constructs into agricultural fields provide unique markers to measure such connectivity. For these reasons, the detection of transgenic DNA in crop landraces is of critical importance. Here we report the presence of introgressed transgenic DNA constructs in native maize landraces grown in remote mountains in Oaxaca, Mexico, part of the Mesoamerican centre of origin and diversification of this crop^{7–9}.

In October and November 2000 we sampled whole cobs of native, or 'criollo', landraces of maize from four standing fields in two locations of the Sierra Norte de Oaxaca in Southern Mexico (samples A1–A3 and B1–B3), more than 20 km from the main mountain-crossing road that connects the cities of Oaxaca and Tuxtpec in the Municipality of Ixtlán. As each kernel results from ovule fertilization by individual pollen grains, each pooled criollo sample represents a composite of ~150–400 pollination events. One additional bulk grain sample (K1) was obtained from the local stores of the Mexican governmental agency Diconsa (formerly the National Commission for Popular Subsistence), which distributes subsidized food throughout the country. Negative controls were cob samples of blue maize from the Cuzco Valley in Peru (P1) and a 20-seed sample from an historical collection obtained in the Sierra Norte de Oaxaca in 1971 (H1). Positive controls were bulk grain

samples of Yieldgard *Bacillus thuringiensis* (Bt)-maize (Bt1; Monsanto Corporation) and Roundup-Ready maize (RR1; Monsanto Corporation) obtained from leftover stock for the 2000 planting season in the United States. Using a polymerase chain reaction (PCR)-based approach, we first tested for the presence of a common element in transgenic constructs currently on the market—the 35S promoter (p-35S) from the cauliflower mosaic virus (CMV). The high copy number and widespread use of p-35S in synthetic vectors used to incorporate transgenic DNA during plant transformation make it an ideal marker to detect transgenic constructs^{10–12}.

We obtained positive PCR amplification using primers specific for p-35S in five of the seven Mexican maize samples tested (Fig. 1). Four criollo samples showed weak albeit clear PCR amplification, whereas the Diconsa sample yielded very strong amplification comparable in intensity to transgenic-positive Bt1 and RR1 controls. The historical negative control (data not shown) and the contemporary sample from Cuzco, Peru, were both invariably negative. Low PCR amplification from landraces was due to low transgenic abundance (that is, a low percentage of kernels in each cob), not to differential efficiency in the reaction, as demonstrated by internal control amplification of the maize-specific alpha zein protein 1 gene (Fig. 1, *zp1*). During the review period of this manuscript, the Mexican Government (National Institute of Ecology, INE, and National Commission of Biodiversity, Conabio) established an independent research effort. Their results, published through official government press releases, confirm the presence of transgenic DNA in landrace genomes in two Mexican states, including Oaxaca. Samples obtained by the Mexican research initiative from sites located near our collection areas in the Sierra Norte de Oaxaca also confirm the relatively low abundance of transgenic DNA in these remote areas. The governmental research effort analysed individual kernels, making it possible for them to quantify abundances in the range of 3–10%. Because we pooled all kernels in each cob, we cannot make such a quantitative statement, although low PCR amplification signal from criollo samples is compatible with abundances in this percentage range.

Using a nested primer system, we were able to amplify the weak bands from all CMV-positive criollo samples (Fig. 1) sufficiently for nucleotide sequencing (GenBank accession numbers AF434747–AF434750), which always showed at least 98% homology with CMV p-35S constructs in commercially used vectors such as pMON273 (GenBank accession number X04879.1) and the K1 sample (accession number AF434746).

Further PCR testing of the same samples showed the presence of the nopaline synthase terminator sequence from *Agrobacterium tumefaciens* (T-NOS) in two of the six criollo samples (A3 and B2; GenBank accession numbers AF434752 and AF434751, respec-

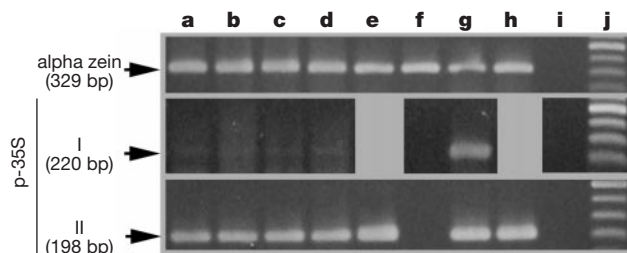


Figure 1 PCR amplification of DNA from the maize-specific alpha zein protein gene (top panel) and the CMV p-35S promoter (centre and bottom panels). The centre panel represents amplification protocol I (single amplification); the bottom panel indicates amplification protocol II (nested priming amplification). **a–d**, Criollo maize samples. Samples A2 (**a**), A3 (**b**), B2 (**c**) and B3 (**d**) are shown. **e**, Sample K1 from Diconsa store. **f**, Negative control P1, from Peru. **g**, Roundup-Ready maize RR1. **h**, Bt-maize Bt1. **i**, Internal negative control for PCR reaction. **j**, DNA ladder (100 base pairs (bp)), 500-bp marker at the top in each panel. Expected size for each fragment is marked on the left.

tively) and the Diconsa sample (K1; accession number AF434753). We detected the *B. thuringiensis* toxin gene *cryIAb* in one criollo sample (B3) (data not shown). We confirmed all of the PCR results through repeated testing.

We performed inverse PCR (iPCR) to reveal the various genomic contexts in which the CMV construct was embedded in the Oaxacan criollo maize. This method enabled us to sequence unknown DNA regions flanking the known p-35S sequence in each of the samples. For each sample, iPCR yielded 1–4 DNA fragments differing in size. We isolated these fragments from electrophoresis gels and attempted to sequence them individually, yielding sequences in eight cases (GenBank accession numbers AF434754–AF434761; Fig. 2). Sequences adjacent to the CMV p-35S DNA were diverse, suggesting that the promoter was inserted into the criollo genome at multiple loci. When compared with GenBank (BLAST, February 2001), two sequences were similar to synthetic constructs containing regions of the *adh1* gene found in transgenic maize currently on the market, such as Novartis Bt11 (Fig. 2, samples A3 and K1). Notably, these two sequences had high homology with each other. Other sequences represented maize-native genomic DNA, including retrotransposon regions, whereas others showed no significant homology with any GenBank sequence (Fig. 2). The diversity of transgenic DNA constructs present in criollo samples suggests the occurrence of multiple introgression events, probably mediated by pollination. In some of these events, the introgressed DNA appeared to have retained its integrity as an unaltered construct (as with *adh1* (ref. 10), whereas in others the transgenic DNA construct seemed to have become re-assorted and introduced into different genomic backgrounds, possibly during transformation or recombination¹³. The apparent predominance of re-assorted sequences obtained in our study might be due to PCR bias for amplification of short fragments, as intact functional constructs are expected to be much longer.

Our results demonstrate that there is a high level of gene flow from industrially produced maize towards populations of progenitor landraces. As our samples originated from remote areas, it is to be expected that more accessible regions will be exposed to higher rates of introgression. Our discovery of a high frequency of transgene insertion into a diversity of genomic contexts indicates that introgression events are relatively common, and that the

transgenic DNA constructs are probably maintained in the population from one generation to the next. The diversity of introgressed DNA in landraces is particularly striking given the existence in Mexico of a moratorium on the planting of transgenic maize since 1998. Whether the presence of these transgenes in 2000 is due to loose implementation of this moratorium, or to introgression before 1998 followed by the survival of transgenes in the population, remains to be established. The intentional release of large amounts of commercial transgenic seed into the environment since the mid-1990s represents a unique opportunity to trace the flow of genetic material over biogeographical regions, as well as a major influence on the future genetics of the global food system.

Further study of the impact of the gene flow from commercial hybrids to traditional landraces in the centres of origin and diversity of crop plants needs to be carefully considered with respect to the future of sustainable food production. Long-term studies should establish whether, or for how long, the integrity of the transgenic construct is retained, and whether the relatively low abundance of transgene introgression detected in the 2000 harvest cycle in Oaxaca will increase, decrease, or remain stable over time. □

Methods

Extraction and purification of genomic DNA

For each cob (sample), all kernels (152–384 kernels per cob) were ground to a fine powder using a steel miller to obtain a pooled sample. Three hundred seeds were also ground from each of the bulk samples, except for the historical negative control, which consisted of 20 seeds. Before use with each sample, millers were thoroughly washed, soaked in 10% sodium perchlorate for 30 min, rinsed and then autoclaved. Genomic DNA was extracted from 100 mg of the powder as described elsewhere¹⁰, with an added purification step using a GeneClean I Kit (Bio 101).

Polymerase chain reaction

For protocol I, amplification reactions using 50–100 ng of extracted genomic DNA were carried out in 25 µl containing 1× PCR buffer (Promega), 2.5 mM MgCl₂, 0.2 mM of each dNTP, 0.5 µM of each primer and 0.625 U of Platinum Taq Polymerase (GibcoBRL). We used a water negative control to verify that reactions were free of contamination. Amplifications were performed on a PTC-100 thermal cycler (MJ Research) with the following parameters: initial denaturation at 95 °C for 2 min; 40 cycles each with denaturing at 95 °C for 45 s, annealing for 1 min at 60 °C/56 °C for CMV/NOS, respectively, extension at 72 °C for 1 min; and a final extension for 5 min at 72 °C. For protocol II, where low amplicon yields were obtained, amplification was repeated as in protocol I but with only 20–25 cycles, followed by a re-amplification or nested amplification of a 1:250 dilution of the PCR products in a new reaction mix with partially or wholly nested primers for 10–15 cycles. Primers cm01 (5'-CACTACAAATGCCATCAT TGCGATA-3') and cm02 (5'-CTTATATAGAGGAAGGGTCTTGCGA-3')¹⁰ were used to detect CMV 35S promoters with protocol I. With protocol II we used nested primer pairs mp3 (5'-TCATCCCTTAC GTCAGTGGAGATAT-3') and mp4 (5'-GATAAAGGAAAGGCCATCGTTGAAG-3')¹², and cm02 and mp4. Primers cm01 and cr02 (5'-CTCTCGCGTAGATTGGTACA-3')¹⁰ were used to detect the *cryIAb* synthetic gene. Primers zp01 (5'-TGCTTGCAATTGTTCCG TCTCTAG-3') and zp02 (5'-GTCGCAGT GACATTGTGGCAT-3')¹⁰ were used to amplify the maize-specific alpha zein protein 1 gene (*zp1*) as an external control for the presence of maize DNA and the efficiency of the reaction.

Inverse PCR

Inverse PCR reactions were modified from previously described protocols^{14,15}. Genomic DNA for iPCR was digested with *EcoRV* (Promega), which targets a single digestion site internal to the p-35S. Restriction fragments were self-ligated with T4 DNA Ligase (Promega) (14 °C, 18 h), followed by heat inactivation (75 °C, 15 min) and phenol extraction. The purified, circularized products were resuspended in 10 µl 1:10 TE (10 mmol l⁻¹ Tris, pH 8.5; 1 mmol l⁻¹ EDTA). PCR amplification was performed using primers designed specifically for the CMV 35S promoter iCMV1 (5'-ACGCTTCAAA GCAAGTGA-3'), iCMV2 (5'-AGTGACAGATAGGATCGGGAAT-3') or iCMV3 (5'-GGAGAGGACACGCTGAAATC-3'), iCMV4 (5'-TAGTGGGATTGTGCGTCATC-3'). These primer pairs were designed to amplify outwards of the 35S promoter, downstream and upstream, respectively.

Nucleotide sequencing

All nucleotide sequencing was carried out at the University of California at San Francisco Comprehensive Cancer Center, Genome Analysis Core Facility. All sequences mentioned above are available on the NCBI GenBank server (<http://www.ncbi.nlm.nih.gov>).

Received 26 July; accepted 31 October 2001.

1. Losey, J. E., Raynor, L. S. & Carter, M. E. Transgenic pollen harms monarch larvae. *Nature* **399**, 214 (1999).

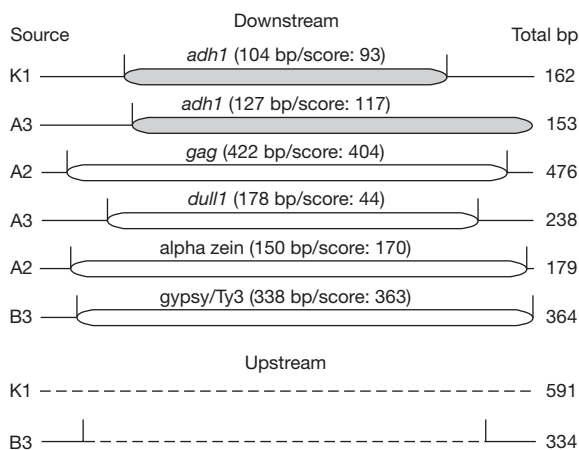


Figure 2 Inverse PCR sequences of flanking regions adjacent to the CMV p-35S in landraces (A2, A3 and B3) and in Diconsa seed (K1). The use of inverse primers results in the appearance of flanking regions as contained within CMV priming sites. For each sequence, best BLAST score is shown in parentheses after the number of nucleotides homologous to the assigned sequence. Vertical lines, end of known CMV sequence; shaded boxes, sequences believed to be derived from transgenic constructs; empty boxes, sequences believed to be native to the maize host; dashed lines, sequences with no homology in GenBank (see Supplementary Information for further details).

2. Saxena, D., Flores, S. & Stotzky, G. Insecticidal toxin in root exudates from *Bt* corn. *Nature* **402**, 480 (1999).
3. Ellstrand, N. C. When transgenes wander, should we worry? *Plant Physiol.* **125**, 1543–1545 (2001).
4. Doebley, J. Molecular evidence for gene flow among *Zea* species—genes transformed into maize through genetic engineering could be transferred to its wild relatives, the Teosintes. *Bioscience* **40**, 443–448 (1990).
5. Ellstrand, N. C., Prentice, H. C. & Hancock, J. F. Gene flow and introgression from domesticated plants into their wild relatives. *Annu. Rev. Ecol. Syst.* **30**, 539–563 (1999).
6. White, S. & Doebley, J. Of genes and genomes and the origin of maize. *Trends Genet.* **14**, 327–332 (1998).
7. Wang, R.-L., Stec, A., Hey, J., Lukens, L. & Doebley, J. The limits of selection during maize domestication. *Nature* **398**, 236–239 (1999).
8. Piperno, D. R. & Flannery, K. V. The earliest archaeological maize (*Zea mays* L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. *Proc. Natl Acad. Sci. USA* **98**, 2101–2103 (2001).
9. Ilitis, H. From teosinte to maize: the catastrophic sexual transmutation. *Science* **222**, 886–894 (1983).
10. Matsuoka, T. *et al.* A method of detecting recombinant DNAs from four lines of genetically modified maize. *Shokuhin Eiseigaku Zasshi* **41**, 137–143 (2000).
11. Gachet, E., Martin, G. G., Vigeau, F. & Meyer, G. Detection of genetically modified organisms (GMOs) by PCR: a brief review of methodologies available. *Trends Food Sci. Technol.* **9**, 380–388 (1999).
12. Anonymous Development of Methods to Identify Foods Produced by Means of Genetic Engineering EU Project SMT4-CT96-2072 (Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin, Berlin, 1999).
13. Pawlowski, W. P. & Somers, D. A. Transgenic DNA integrated into the oat genome is frequently interspersed by host DNA. *Proc. Natl Acad. Sci. USA* **95**, 12106–12110 (1998).
14. Hartl, D. L. & Ochman, H. in *Methods in Molecular Biology* (ed. Harwood, A.) 293–301 (Humana, Totowa, New Jersey, 1996).
15. Zimmermann, A., Lüthy, L. & Pauli, U. Event specific transgene detection in Bt11 corn by quantitative PCR at the integration site. *Lebensm.-Wiss. Technol.* **33**, 210–216 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank the Union de Comunidades Zapoteco Chinanteca (UZACHI) for access to their field laboratory, Y. Lara (Estudios Rurales y Asesoría, Oaxaca) for facilitation, A. King for Peruvian maize samples and CIMMYT maize germplasm bank for the historical control.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.H.C. (e-mail: ichapela@nature.berkeley.edu).

Kranz anatomy is not essential for terrestrial C₄ plant photosynthesis

Elena V. Voznesenskaya*, Vincent R. Franceschi†, Olavi Kilrats†, Helmut Freitag‡ & Gerald E. Edwards†

* Laboratory of Anatomy and Morphology, V. L. Komarov Botanical Institute of Russian Academy of Sciences, Prof. Popov Street 2, 197376 St Petersburg, Russia

† School of Biological Sciences, Washington State University, Pullman, Washington 99164-4236, USA

‡ Morphologie u. Systematik der Pflanzen, Universität Gesamthochschule Kassel, FB19, D-34109 Kassel, Germany

An important adaptation to CO₂-limited photosynthesis in cyanobacteria, algae and some plants was development of CO₂-concentrating mechanisms (CCM)¹. Evolution of a CCM occurred many times in flowering plants, beginning at least 15–20 million years ago, in response to atmospheric CO₂ reduction, climate change, geological trends, and evolutionary diversification of species². In plants, this is achieved through a biochemical inorganic carbon pump called C₄ photosynthesis, discovered 35 years ago³. C₄ photosynthesis is advantageous when limitations on carbon acquisition are imposed by high temperature, drought and saline conditions. It has been thought that a specialized leaf anatomy, composed of two, distinctive photosynthetic cell types (Kranz anatomy), is required for C₄ photosynthesis⁴. We provide

evidence that C₄ photosynthesis can function within a single photosynthetic cell in terrestrial plants. *Borszczowia aralocaspica* (Chenopodiaceae) has the photosynthetic features of C₄ plants, yet lacks Kranz anatomy. This species accomplishes C₄ photosynthesis through spatial compartmentation of photosynthetic enzymes, and by separation of two types of chloroplasts and other organelles in distinct positions within the chlorenchyma cell cytoplasm.

CO₂-concentrating mechanisms (CCM) have evolved that increase the level of CO₂ at the site of fixation by the C₃ photosynthetic pathway via the enzyme ribulose 1,5-bisphosphate (RuBP) carboxylase/oxygenase (Rubisco); these CCMs also negate the counterproductive oxygenase activity of Rubisco. Plants that lack a CCM directly fix atmospheric CO₂ in their photosynthetic cells via Rubisco. They are called C₃ plants⁵ because the initial product of fixation is a three-carbon compound, and they show high rates of photorespiration owing to the oxygenase activity of Rubisco. CCM in terrestrial plants occur via a C₄ dicarboxylic acid pathway; thus they are called C₄ plants. C₄ plants actively take up CO₂ from the atmosphere and concentrate it around Rubisco for assimilation into organic matter. This requires spatial separation of fixation of atmospheric CO₂ (via phosphoenolpyruvate carboxylase) into C₄ acids, and donation of CO₂ from C₄ acids (via C₄ acid decarboxylases) to RuBP carboxylase of the C₃ pathway.

Photosynthesis has been thought to occur in all terrestrial C₄ plants by the cooperative function of two types of photosynthetic tissue: an inner layer called Kranz or bundle sheath cells, and an outer layer of palisade cells^{4,6}. Whereas the CCM in terrestrial plants occur via a C₄ dicarboxylic acid pathway, cyanobacteria and algae employ different mechanisms¹. Besides C₃ and C₄ plants, some vascular plants fix atmospheric CO₂ at night through a C₄ pathway and further process the carbon via the C₃ pathway during the day (called crassulacean acid metabolism or CAM)⁷. This results in a temporal separation of the process rather than a spatial separation, such as in Kranz anatomy, as occurs in C₄ plants. The process of photosynthesis in C₃ and CAM plants is achieved within a single photosynthetic cell, without Kranz anatomy.

Our evidence that Kranz anatomy is not essential for C₄ plant photosynthesis in terrestrial species is based on studies with the monotypic genus *Borszczowia*. *Borszczowia aralocaspica* Bunge (subfamily Salsoloideae, family Chenopodiaceae) grows in salty depressions of Central Asian semi-deserts. It is a succulent species with unusual chlorenchyma, and its carbon isotope composition is like that of C₄ or obligate CAM plants⁸. The Chenopodiaceae family has the largest number of C₄ species among dicotyledonous plants⁴; it has high diversity in evolution of C₄ photosynthesis, including five variants of Kranz anatomy^{8,9} and two variants of C₄ biochemistry^{10–12}.

Figure 1 shows the leaf anatomy of *B. aralocaspica*, *Salsola laricina* and *Suaeda heterophylla*, all in subfamily Salsoloideae, family Chenopodiaceae. The general leaf anatomy of *B. aralocaspica* (Fig. 1a), looks similar to the C₄ salsoloid type as demonstrated by the C₄ plant *S. laricina* Pall (Fig. 1b)¹⁰. *S. laricina* has a central, main vein surrounded by water storage parenchyma, and Kranz anatomy with distinctive peripheral layers of palisade and Kranz cells. The Kranz cell chloroplasts have grana and accumulate starch, whereas the palisade chloroplasts have reduced grana and lack starch¹⁰.

However, in contrast, *B. aralocaspica* has only a single layer of unusual palisade-shaped chlorenchyma cells, which are located between the central water storage tissue and the hypodermal cells (Fig. 1a, also see ref. 8). These radially elongated chlorenchyma cells have a large, central vacuole and a layer of peripheral cytoplasm with few chloroplasts in the distal (from the vascular bundle) part of the cell and a high density of cytoplasm with numerous chloroplasts and large mitochondria (the latter observed by electron microscopy) in the proximal position (see arrows in Fig. 1a). Chloroplasts

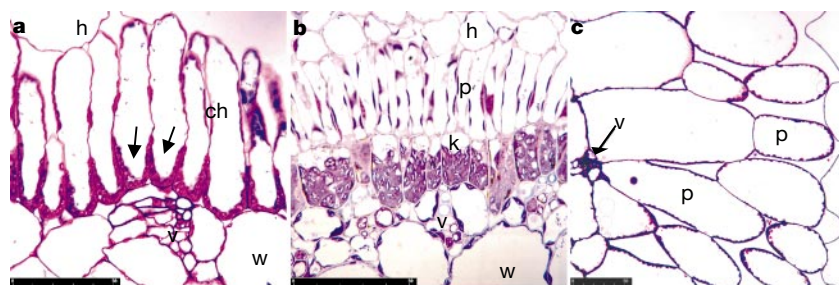


Figure 1 Light microscopy of leaves (transverse sections). **a**, *Borszczowia aralocaspica*; **b**, *Salsola laricina*; and **c**, *Suaeda heterophylla*. ch, chlorenchyma cell; h, hypodermal cell; k, Kranz cell; p, palisade cell; v, vascular tissue; w, water storage cell. Scale bars, 50 μ m.

in the distal part of the cell lack grana, and are without starch, while those in the proximal part have grana and contain starch (Fig. 1a, also see ref. 8; observations on degree of grana formation were made by electron microscopy, not shown). There are intercellular air spaces between the distal parts, but not the proximal parts, of the chlorenchyma cells.

Thus, this plant has leaf anatomy similar to C_4 *Salsola*, with one notable exception: it has dimorphic chloroplasts in a single photosynthetic cell instead of in two cell types with Kranz anatomy. For comparison, Fig. 1c shows the leaf anatomy of *Su. heterophylla* (Kar. & Kir.) Bunge, a C_3 species that is also in subfamily Salsoloideae. It lacks Kranz anatomy and has two to three layers of large, palisade-

like mesophyll cells with a large central vacuole and a thin layer of cytoplasm with chloroplasts along the cell periphery, which is common for C_3 photosynthetic cells.

The unique cytological features of *B. aralocaspica* suggest that its photosynthetic cells may be biochemically compartmentalized for carbon assimilation, and represent a new photosynthesis mechanism. We have tested this hypothesis with a number of corroborative techniques. Using immunolocalization techniques, we have demonstrated the compartmentation of important photosynthetic enzymes in the chlorenchyma cells. Rubisco is concentrated in chloroplasts in the proximal part of the cell (Fig. 2a and c). With respect to enzymes of the C_4 pathway, PEP carboxylase is abundant

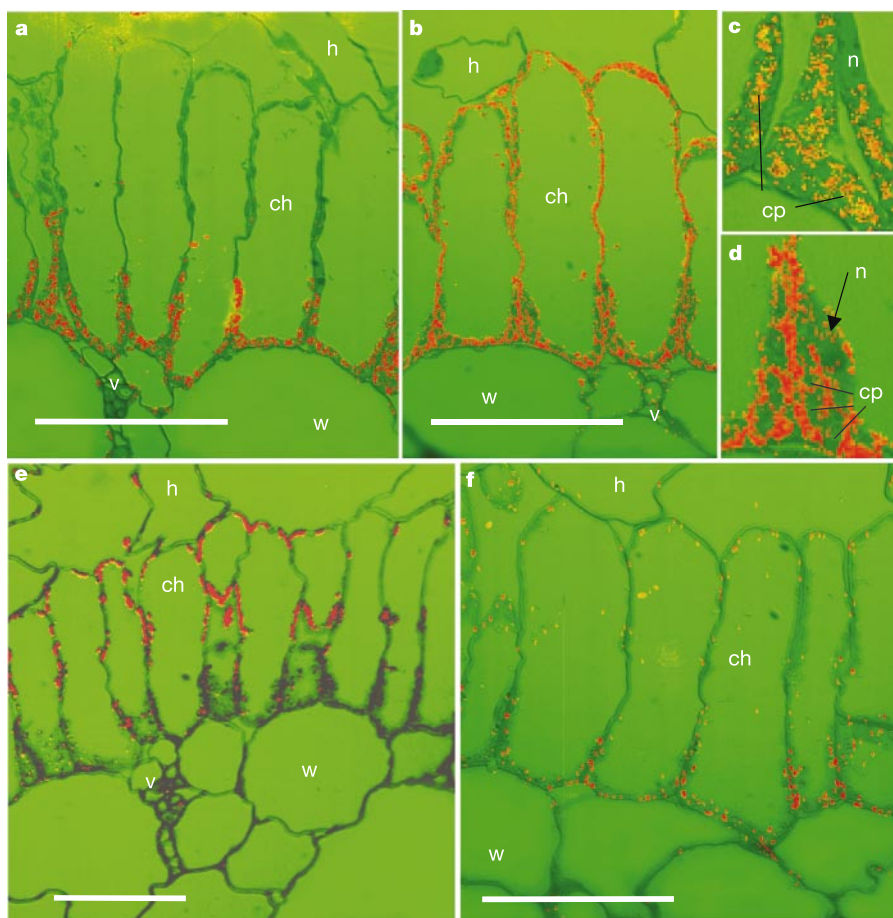


Figure 2 Immunolocalization of photosynthetic enzymes in leaves of *Borszczowia aralocaspica* by confocal laser scanning microscopy. Immunolocalization of Rubisco (**a**), PEP carboxylase (**b**), higher magnification showing Rubisco in chloroplasts in proximal end of cell (**c**), higher magnification showing PEP carboxylase in cytosol (**d**), pyruvate, Pi

dikinase (**e**) and NAD-malic enzyme (**f**). Red dots indicate where the enzyme is present. cp, chloroplast; ch, chlorenchyma cell; h, hypodermal cell; n, nucleus; v, vascular tissue; w, water storage cell. Scale bars, 50 μ m.

Table 1 Activities of photosynthetic enzymes

Species	Enzyme* activity ($\mu\text{mol mg chlorophyll}^{-1} \text{ min}^{-1}$)				
	Rubisco	PPDK	PEPC	NAD-ME	NADP-ME
<i>Suaeda heterophylla</i> (C ₃)	7.07 $\pm$ 0.33	0.25	0.3 $\pm$ 0.03	2.80 $\pm$ 0.45	n.d.
<i>Salsola laricina</i> (C ₄)	1.13	10.33	20.5	12.7 $\pm$ 3.1	0.39 $\pm$ 0.12
<i>Borszczowia aralocaspica</i>	2.17 $\pm$ 0.15	8.52 $\pm$ 0.42	12.8 $\pm$ 0.5	3.76 $\pm$ 0.45	2.42 $\pm$ 0.53

* PPDK, pyruvate, Pi dikinase; PEPC, PEP carboxylase; NAD-ME, NAD-malic enzyme; NADP-ME, NADP-malic enzyme.

Temperature of assay was 25 °C. n.d., not detected. For methods for assay of enzymes see refs 12 and 24. Standard errors where shown were for two independent replications.

throughout the cytosol (Fig. 2b and d), and pyruvate, Pi dikinase (PPDK) is located in chloroplasts in the distal part of the cell (Fig. 2e), whereas NAD-ME (NAD-malic enzyme) is located proximally in the cell (Fig. 2f) in mitochondria (immunogold labelling by electron microscopy, not shown). Water storage and hypodermal cells have a few, small chloroplasts that show labelling for Rubisco, but not for enzymes of C₄ photosynthesis. This indicates the succulent water storage cells are not performing CAM⁷.

Borszczowia aralocaspica has activities of photosynthetic enzymes that are characteristic of C₄ plants¹³, as seen by comparison with the C₄ *S. laricina* (NAD-ME subtype) (Table 1). The activities of enzymes of the C₄ cycle, PPDK, PEP carboxylase, and malic enzyme, are high and more than sufficient to support the measured rates of CO₂ assimilation (Fig. 3). As we already noted, the immunolocalization results show that NAD-ME is located proximally in the mitochondria in photosynthetic cells in *B. aralocaspica*.

The mitochondrial NAD-ME may react with both NAD⁺ and (to a lesser extent) NADP⁺ as substrate (Table 1), because no NADP-ME protein was detected on western blots or by immunolocalization (using antibody of NADP-ME from maize). In comparison, *Su. heterophylla* has activities of Rubisco, PEP carboxylase, PPDK and NADP-ME characteristic of C₃ species. It has significant activity of NAD-ME, a constitutive mitochondrial enzyme that is readily detected in C₃ tissue^{12,14}.

The carbon isotope composition ($\delta^{13}\text{C}$) of the three species is given in Table 2. This diagnostic feature can be used to distinguish types of photosynthetic mechanisms based on the carboxylation enzymes operating. C₃ plants have a more negative value of $\delta^{13}\text{C}$ because atmospheric CO₂ is fixed by Rubisco, which discriminates against ¹³CO₂. C₄ plants have a more positive value because atmospheric CO₂ is fixed by PEP carboxylase, which does not discriminate against ¹³CO₂. The CO₂, which is subsequently donated to Rubisco, is compartmentalized and largely fixed with minimal leakage¹⁵. The high $\delta^{13}\text{C}$ value for *S. laricina* is typical of that of C₄ plants, while the low value for *Su. heterophylla* is typical of that for C₃ plants. Of significance to the hypothesis of operation of a single-cell C₄ mechanism, *B. aralocaspica* has $\delta^{13}\text{C}$ values that are clearly of C₄ origin.

Rubisco is a bifunctional enzyme, where CO₂ and O₂ are competitive substrates for reaction with RuBP¹⁶. Whereas reaction of RuBP with CO₂ results in carbon assimilation, reaction with O₂ results in photorespiration. This is an apparently counterproductive process that is a significant component of photosynthetic activity in C₃ plants, but not in C₄ plants, owing to the localized CO₂-concentrating effect of the C₄ mechanism. Thus, the response of photosynthesis to varying CO₂ and O₂ is also diagnostic for the C₃ versus C₄ mechanism of carbon assimilation. Ambient levels of O₂ (21%) inhibit photosynthesis in *Su. heterophylla* at varying levels of CO₂, a characteristic of C₃ plant photosynthesis, whereas photosynthesis in *S. laricina* is insensitive to O₂, as is typical of C₄ plants (Fig. 3a, b). The response of *B. aralocaspica* is also like that of C₄ plants, in that O₂ does not inhibit photosynthesis (Fig. 3c); in fact, there is some stimulation by O₂ under limiting CO₂, which has been previously observed in some C₄ species¹⁷. Even at the CO₂ compensation point, where there is no net CO₂ fixation, O₂ does not affect photosynthesis. This suggests that photorespiration is restricted, and that, at this point, the rate of CO₂ fixation by Rubisco equals the rate of dark respiration. *B. aralocaspica* also has substantial rates of respiration in the dark, ruling out the possibility that its growth occurs through fixation of atmospheric CO₂ in the dark through PEP carboxylase via crassulacean acid metabolism (CAM).

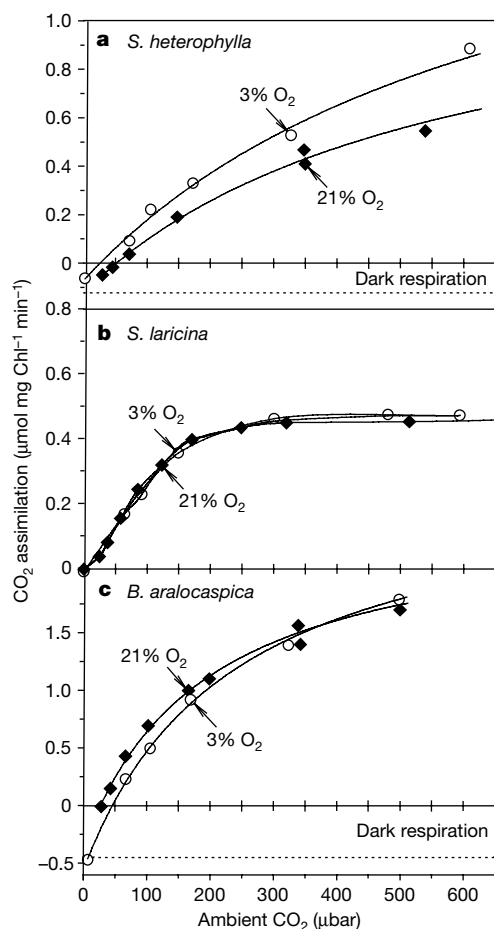


Figure 3 The response of photosynthesis to varying atmospheric levels of CO₂. **a**, *Suaeda heterophylla* (C₃); **b**, *Salsola laricina* (C₄); and **c**, *Borszczowia aralocaspica*. The rates of photosynthesis were measured as described²⁶. Conditions were 25 °C, a light intensity of 1,200 $\mu\text{mol quanta m}^{-2} \text{ s}^{-1}$, and varying levels of CO₂ (C_a) and O₂, as shown.

Table 2 Comparison of carbon isotope composition ($\delta^{13}\text{C}$) of leaf tissue

Species	Carbon isotope value ($\delta^{13}\text{C}$)
<i>Suaeda heterophylla</i> (C ₃)	-25.34*, -27.28†
<i>Salsola laricina</i> (C ₄)	-14.80‡
<i>Borszczowia aralocaspica</i>	-12.37*, -13.78§

The $\delta^{13}\text{C}$ isotope values were determined as described²⁵.

* Original data on leaves from plant material used in current study.

† Original data from leaves kindly provided by W. Stichler, average of two replications.

‡ Data from ref. 12.

§ Data from ref. 8.

Thus our results prove that C_4 photosynthesis exists in a terrestrial plant without the dual-cell Kranz anatomy system. C_4 photosynthesis is accomplished within a single cell by novel cytological features that allow spatial separation of the biochemical events necessary for operation of the C_4 mechanism. Until now, the separation of functions in terrestrial C_4 plants has been associated only with Kranz-type leaf anatomy. *B. aralocaspica* has evolved a unique solution for the requirement of spatial separation of these biochemical functions within a single cell. PPDK is positioned at the distal part of the cell, where it can generate PEP, the substrate for PEP carboxylase. PEP carboxylase fixes atmospheric CO_2 supplied to the cell through the adjoining intercellular air spaces. NAD-ME and Rubisco are compartmentalized to the interior of the cells, where CO_2 can be donated from C_4 acids to the C_3 pathway. The C_4 -type $\delta^{13}C$ values and lack of inhibition of photosynthesis by O_2 demonstrate that CO_2 can be concentrated sufficiently around Rubisco through this specialized compartmentation to minimize photorespiration. The physical requirements for C_4 photosynthesis may be met by the existence of a sufficiently high diffusive resistance in the aqueous phase between sites of CO_2 donation to Rubisco in the proximal ends of the cells and sites of fixation of atmospheric CO_2 by PEP carboxylase at the distal ends.

Our results are relevant to the discussion of evolution of C_4 photosynthesis in plants, because the first land plants were C_3 species^{2,4}. C_4 species are an important component of global ecosystems and there is interest in their evolution and the consequences to evolution of mammals^{2,4,18,19}. Palaeorecords have been used to study how long C_4 plants have existed on Earth by finding well preserved fossils that have Kranz anatomy and C_4 -type isotope composition. Now our results indicate that it is possible that plant fossils with a C_4 isotope composition but without Kranz anatomy may be C_4 species (rather than CAM species).

C_4 plants are also of considerable interest because this mechanism of photosynthesis has an advantage over C_3 plants for conversion of solar energy into biomass in hot, dry and/or saline habitats. Maize, sugarcane and sorghum are important C_4 crop plants, but most agricultural crops, including rice and wheat, are C_3 plants. This has led to interest in genetically engineering C_3 crops to perform C_4 photosynthesis in order to increase productivity^{20,21}. Although this may require alterations in anatomy as well as biochemistry, our results indicate that it would not require development of two photosynthetic cell types. □

Methods

Plants were grown under controlled growth conditions with day/night temperatures of 25/18 °C, and a 14/10 h photoperiod, with a stepwise increase and decrease in light intensity during the day to a maximum photosynthetic quantum flux density of 1,100 $\mu mol m^{-2} s^{-1}$. For immunolocalization studies, samples were prepared as described¹⁰. Antibodies used were anti-spinach Rubisco (LSU) IgG (courtesy of B. McFadden), anti-maize PPDK IgG (courtesy of T. Sugiyama), anti-maize PEPCK IgG (Chemicon), anti-maize NADP-ME IgG with relative molecular mass (M_r) 62,000 (courtesy of C. Andreo²²), and anti-*Amaranthus hypochondriacus* mitochondrial NAD-ME IgG, which was prepared against the α subunit with M_r 65,000 (courtesy of J. Berry²³). For details of techniques used for immunolocalization by light microscopy see ref. 10. The background labelling with preimmune serum was non-specific and low to nonexistent (results not shown but see ref. 10).

Received 14 August; accepted 18 October 2001.

1. Badger, M. R. & Spalding, M. H. in *Photosynthesis: Physiology and Metabolism* (eds Leegood, R. C., Sharkey, T. D. & von Caemmerer, S.) 369–397 (Kluwer Academic, Netherlands, 2000).
2. Sage, R. F. Environmental and evolutionary preconditions for the origin and diversification of the C_4 photosynthetic syndrome. *Plant Biol.* **3**, 202–213 (2001).
3. Hatch, M. D. in *Photosynthesis and Photorespiration* (eds Hatch, M. D., Osmond, C. B. & Slatyer, R. O.) 139–152 (Wiley-Interscience, New York, 1971).
4. Sage, R. F. & Monson, R. K. *C₄ Plant Biology* (Academic, San Diego, 1999).
5. Edwards, G. E. & Walker, D. A. *C₃/C₄ Mechanisms, and Cellular and Environmental Regulation, of Photosynthesis* (Blackwell Scientific, Oxford, 1983).
6. Edwards, G. E., Furbank, R. T., Hatch, M. D. & Osmond, C. B. What does it take to be C_4 ? Lessons from the evolution of C_4 photosynthesis. *Plant Physiol.* **125**, 46–49 (2001).
7. Winter, K. & Smith, J. A. C. *Crassulacean Acid Metabolism* (Springer, New York, 1996).
8. Freitag, H. & Sticher, W. A remarkable new leaf type with unusual photosynthetic tissue in a central asiatic genus of Chenopodiaceae. *Plant Biol.* **2**, 154–160 (2000).

9. Carolin, R. C., Jacobs, S. W. L. & Vesk, M. The structure of the cells of the mesophyll and parenchymatous bundle sheath of the Gramineae. *Bot. J. Linn. Soc.* **66**, 259–275 (1973).
10. Voznesenskaya, E. V., Franceschi, V. R., Pyankov, V. I. & Edwards, G. E. Anatomy, chloroplast structure and compartmentation of enzymes relative to photosynthetic mechanisms in leaves and cotyledons of species in the tribe Salsola (Chenopodiaceae). *J. Exp. Bot.* **50**, 1779–1795 (1999).
11. Voznesenskaya, E. V. & Gamale, Y. V. The ultrastructural characteristics of leaf types with Kranz-anatomy. *Bot. Zh.* **71**, 1291–1307 (1986) (in Russian).
12. Pyankov, V. I. et al. Occurrence of C_3 and C_4 photosynthesis in cotyledons and leaves of *Salsola* species (Chenopodiaceae). *Photosynth. Res.* **63**, 69–84 (2000).
13. Kanai, R. & Edwards, G. in *C₄ Plant Biology* (eds Sage, R. F. & Monson, R. K.) 49–87 (Physiological Ecology series, Academic, San Diego, 1999).
14. Pyankov, V. I. et al. Features of photosynthesis in *Haloxylon* species of Chenopodiaceae that are dominant plants in Central Asian deserts. *Plant Cell Physiol.* **40**, 125–134 (1999).
15. Farquhar, G. D. On the nature of carbon isotope discrimination in C_4 species. *Aust. J. Plant Physiol.* **10**, 205–226 (1983).
16. Laing, W. A., Ogren, W. L. & Hageman, R. H. Regulation of soybean net photosynthetic CO_2 fixation by the interaction of CO_2 , O_2 , and ribulose 1,5-diphosphate carboxylase. *Plant Physiol.* **54**, 678–685 (1974).
17. Maroco, J. P., Ku, M. S. B. & Edwards, G. E. Utilization of O_2 in the metabolic optimization of C_4 photosynthesis. *Plant Cell Environ.* **23**, 115–121 (2000).
18. Cerling, T. E., Quade, J. & Wang, Y. Expansion and emergence of C_4 plants. *Nature* **371**, 112 (1994).
19. Pagani, M., Freeman, K. H. & Arthur, M. A. Late Miocene atmospheric CO_2 concentrations and the expansion of C_4 grasses. *Science* **285**, 876–878 (1999).
20. Sheehy, J. E., Mitchell, P. L. & Hardy, B. *Redesigning Rice Photosynthesis to Increase Yield* (IRRI and Elsevier Science, Makati City, Philippines, 2000).
21. Matsuoka, M., Furbank, R. T., Fukayama, H. & Miyao, M. Molecular engineering of C_4 photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 297–314 (2001).
22. Maurino, V. G., Drincovich, M. F. & Andreo, C. S. NADP-malic enzyme isoforms in maize leaves. *Biochem. Mol. Biol. Int.* **38**, 239–250 (1996).
23. Long, J. J., Wang, J.-L. & Berry, J. O. Cloning and analysis of the C_4 photosynthetic NAD-dependent malic enzyme of amaranth mitochondria. *J. Biol. Chem.* **269**, 2827–2833 (1994).
24. Voznesenskaya, E. V. et al. *Salsola arbusculiformis*, a C_3 – C_4 intermediate in Salsola (Chenopodiaceae). *Ann. Bot.* **88**, 337–348 (2001).
25. Akhiani, H., Trimborn, P. & Ziegler, H. Photosynthetic pathways in *Chenopodiaceae* from Africa, Asia and Europe with their ecological, phytogeographical and taxonomical importance. *Plant Syst. Evol.* **206**, 187–221 (1997).
26. Dai, Z., Ku, M. S. B. & Edwards, G. E. Oxygen sensitivity of photosynthesis in C_3 , C_4 , and C_3 – C_4 intermediate species of *Flaveria*. *Planta* **198**, 563–571 (1996).

Acknowledgements

We thank the Electron Microscope Center of Washington State University for use of facilities and staff assistance. We thank N. Ogar and H. Akhiani for supply of viable seeds, and S. Huber, C. A. Ryan and T. Okita for comments on the manuscript. This work was supported by the National Science Foundation (G.E.E.) and the DFG (H.F.).

Correspondence and requests for materials should be addressed to G.E. (e-mail: edwards@wsu.edu).

Interactive memory systems in the human brain

R. A. Poldrack[†], J. Clark^{*}, E. J. Paré-Blagojev[†], D. Shohamy[‡], J. Crespo Moyano[‡], C. Myers[§] & M. A. Gluck[‡]

^{*} Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, and Harvard Medical School, Charlestown, Massachusetts 02131, USA

[†] Harvard Graduate School of Education, Cambridge, Massachusetts 02138, USA

[‡] Center for Molecular and Behavioural Neuroscience; and [§] Department of Psychology, Rutgers University, Newark, New Jersey 07102, USA

Learning and memory in humans rely upon several memory systems, which appear to have dissociable brain substrates^{1,2}. A fundamental question concerns whether, and how, these memory systems interact. Here we show using functional magnetic resonance imaging (fMRI) that these memory systems may compete with each other during classification learning in humans. The medial temporal lobe and basal ganglia were differently engaged across subjects during classification learning depending upon whether the task emphasized declarative or nondeclarative memory, even when the to-be-learned material and the level of performance did not differ. Consistent with competition between

memory systems suggested by animal studies^{3,4} and neuro-imaging⁵, activity in these regions was negatively correlated across individuals. Further examination of classification learning using event-related fMRI showed rapid modulation of activity in these regions at the beginning of learning, suggesting that subjects relied upon the medial temporal lobe early in learning. However, this dependence rapidly declined with training, as predicted by previous computational models of associative learning^{6–8}.

In experiment 1, we used fMRI to image brain activity during performance of classification tasks designed to emphasize either declarative or nondeclarative memory processes. Subjects engaged in a category learning ('weather prediction') task with probabilistic cue–outcome relations based on trial-by-trial feedback^{9,10}. This type of task, which involves online learning of stimulus–response associations, is thought to engage nondeclarative memory^{11,12}, and previous research with this task revealed that patients with basal ganglia disorders were impaired^{13,14}. Amnesic patients with damage to the medial temporal lobe (MTL) demonstrated normal learning during the early portion of training but were impaired relative to controls as learning progressed^{10,13}. We compared performance of normal subjects on this feedback-based (FB) version of the weather-prediction task with performance on a version of the task designed to emphasize declarative memory processes (see Fig. 1). Rather than FB learning, subjects learned the stimuli and categories in a paired-associate (PA) manner and were then tested on their classification ability after learning. Paired-associate learning, which requires learning associations between previously unrelated pairings of stimuli, is thought to engage declarative memory, and is known to rely strongly upon the MTL^{15,16}. Classification ability at the end of training was similar for both the FB and PA version of the task (86.5% for FB versus 82.7% for PA; $F_{1,21} = 1.13$, $P = 0.30$), suggesting that subjects in both tasks acquired knowledge capable of supporting comparable accuracy.

During fMRI scanning at 1.5 T in experiment 1, subjects alternated between the classification task and a baseline task with similar perceptual and motor characteristics but no learning demands. The FB and PA versions of the task (given to separate groups of 13 subjects) differed in the timing of category presentation and the nature of the response (see Fig. 1), but the stimuli used across the two versions of the task were identical and both required a button-press response. Statistical parametric mapping was performed to identify regions whose activity was significantly different between classification and baseline tasks (see Fig. 2). Engagement in the FB weather-prediction task resulted in activation of the basal ganglia (caudate nucleus) in addition to widespread activation of cortical regions compared to the baseline task. Conversely, activity during the FB task was significantly below baseline in a number of regions; many of these regions (medial frontal and parietal cortex, auditory cortex) are often deactivated during performance of demanding visual tasks, but in this case the MTL was also deactivated. This pattern of activity is consistent with a previous study of the FB task using a different baseline task⁵.

In order to determine whether MTL deactivation was specific to FB learning, we compared activity during the FB task with that observed during the PA weather-prediction task in a separate group of subjects. Comparison between groups demonstrated significant task-dependent modulation of both the MTL (PA > FB) and the caudate nucleus (PA < FB) (see Fig. 2c). This difference in MTL and caudate responses between FB and PA tasks suggests that the effects of these structures are negatively related. Random-effects functional connectivity analysis was performed to determine more directly whether these regions exhibit negatively correlated activity. We extracted signal change for each subject from the left MTL region showing maximal task-dependent modulation in the PA > FB comparison (stereotactic coordinates [−33, −30, −18], 6 mm radius). We then entered the signal change data into a correlation analysis across subjects with signal change in all

voxels. This analysis identified a right caudate region whose activity was negatively correlated with the left MTL, along with several other regions (see Fig. 2). Although this correlation cannot establish causation, it provides further evidence for a negative relationship between these structures.

These results suggest that engagement of the declarative and nondeclarative memory systems may be modulated by task demands when the material to be learned is identical. To determine whether the engagement of these systems changes dynamically over the course of learning, another group of 14 subjects participated in experiment 2 using event-related fMRI at 3 T with the FB task. The use of event-related fMRI allowed the decomposition of responses to individual trials, and further allowed parametric analysis of changes in the event-related response over the course of learning¹⁷.

Subjects performed 96 classification trials over the course of two fMRI scans. As in experiment 1, the basal ganglia and other cortical and subcortical regions were active compared to baseline, and the MTL was deactivated (see Fig. 3). Examination of parametric changes (specifically, exponential changes) in the evoked response over time demonstrated that MTL was initially active and caudate was initially inactive, but that the MTL quickly became deactivated and the caudate nucleus became activated. We saw no parametric change in the MTL during the second block of trials (that is, the MTL was deactivated throughout), whereas the caudate nucleus did show a renewed dip and subsequent rise in the second block.

On the basis of the results from experiment 2, we re-analysed the data from the FB task in experiment 1 for parametric changes in response, using the regions identified in experiment 2 as a priori regions of interest (8-mm-diameter sphere centred at local maximum). A significant parametric increase was observed in the caudate nucleus region, confirming the rapid changes in caudate activity observed in the event-related study, whereas the parametric decrease in MTL was not significant ($P = 0.114$).

The results presented here provide the first substantive evidence, to our knowledge, for competition between memory systems in the

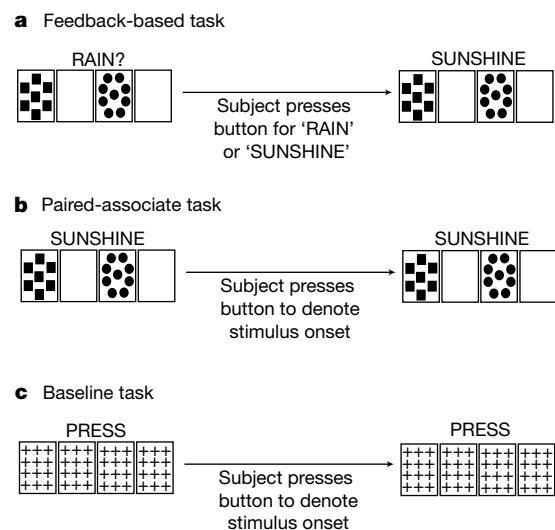


Figure 1 Depiction of task design for experiment 1. **a**, In the feedback-based (FB) task (also used in experiment 2), the stimulus was initially presented with a cue to respond (in this case, 'RAIN'), and the subject responded according to whether he/she thought the particular set of cards was associated with rain or sunshine by pressing one of two keys. The feedback (that is, the outcome for that trial) was then presented after 4 s. **b**, In the paired-associate (PA) task, the outcome was presented initially at the same time as the stimulus, and remained on the screen throughout the trial. Subjects pressed a single key simply to note the appearance of the stimulus. **c**, In the baseline task, subjects were presented with a set of crosshair stimuli (different from those used in the classification trials) and pressed a single key simply to note the appearance of the stimulus.

human brain. Whereas our previous work demonstrated MTL deactivation and striatal activation during classification learning⁵, the present study provides direct evidence for competition at the neural level by demonstrating three essential features of the MTL–striatum interaction. First, it shows that engagement of MTL and striatum is modulated by whether the task encourages the use of declarative versus nondeclarative memory processes or strategies. Second, it demonstrates that engagement of these regions is negatively correlated across subjects. Third, it demonstrates rapid reciprocal changes in the engagement of these regions. These data are concordant with animal lesion studies demonstrating that the memory systems based on the MTL and striatum can compete with one another during learning^{3,4,18}.

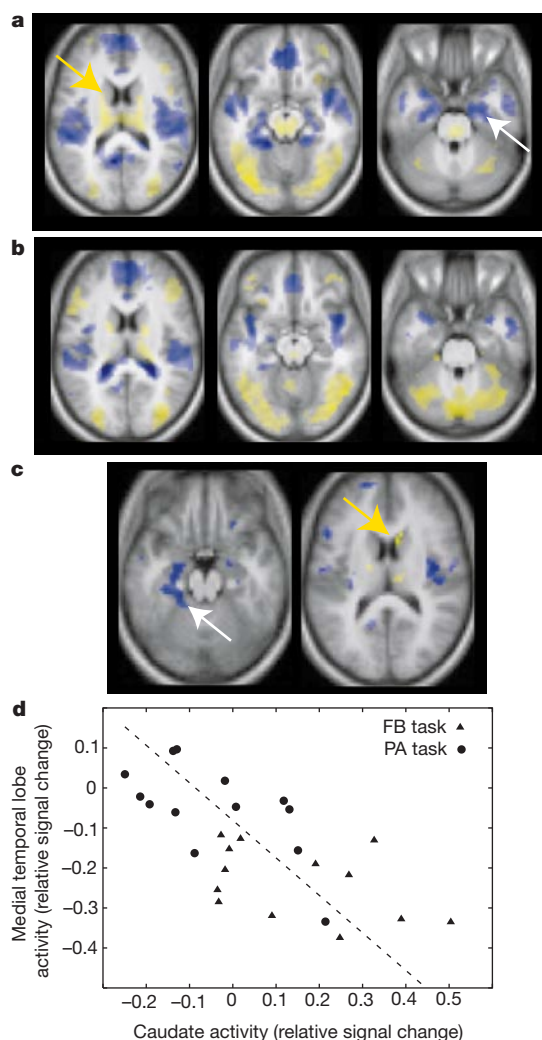


Figure 2 Results from FB versus PA classification learning (experiment 1). **a**, Activation for FB compared to baseline; areas in yellow were more active during weather prediction compared to baseline, whereas regions in blue were less active during weather prediction than baseline ($P < 0.005$ uncorrected, 5 voxel extent threshold). Yellow arrow highlights region of caudate activation, white arrow highlights region of MTL deactivation.

b, Activation for PA compared to baseline. **c**, Regions exhibiting significant differences between FB and PA classification learning tasks. (yellow: FB > PA, blue: PA > FB). Yellow arrow highlights caudate region with FB > PA, white arrow highlights MTL region with PA > FB. **d**, Plot of task-related signal change from the MTL region exhibiting maximal task-dependent differences (centred at $[-33, -30, -18]$, with a 6 mm radius) against a region in the right caudate that exhibited significant negative correlation with the MTL in functional connectivity analysis (centred at $[9, 6, 21]$, with a 6 mm radius). Each data point represents a single subject.

Previous studies of classification learning found that patients with MTL lesions were impaired after several hundred training trials compared to matched controls^{10,13}. These patients did not exhibit significant deficits in early training (through 50 trials), although subsequent studies have found early-learning deficits as well¹⁹. Patients with basal ganglia disorders (Parkinson's and Huntington's diseases) were severely impaired on the task from the onset of training and did not learn the task even after many training trials^{13,14}. These findings have been taken to suggest that category learning relies primarily upon the striatum but that the MTL becomes important later in training^{10,13}. The results of the present study, however, show that the MTL is engaged very early in learning and that it then becomes deactivated throughout training (up to 144 trials in experiment 1). Because the task used in the fMRI studies reported here differs in several ways from the version used in the

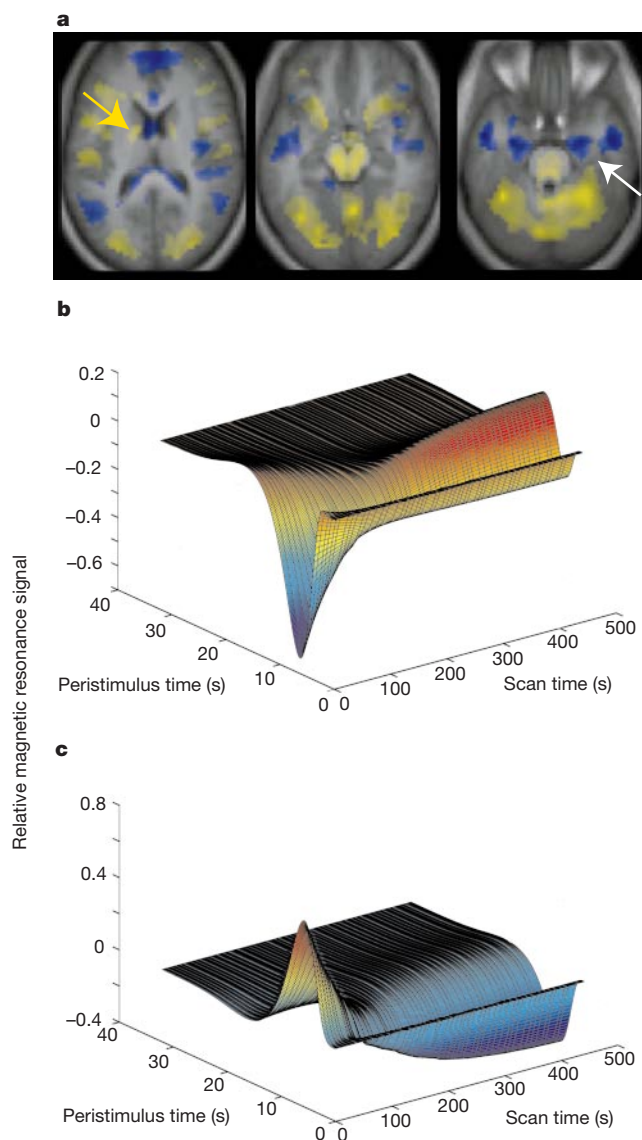


Figure 3 Results from event-related fMRI study of FB category learning (experiment 2).

a, Regions exhibiting significant evoked activation (yellow) or deactivation (blue) for classification trials. Yellow arrow highlights region of caudate activation, white arrow highlights region of MTL deactivation. **b**, **c**, Depiction of parametric change in modelled evoked haemodynamic response across the initial 450-s scanning run (averaged across subjects) in **b**, left body of caudate nucleus ($-12, 3, 21$), and **c**, left MTL ($-24, -3, -24$). Red indicates positive, event-related response, blue indicates negative event-related response.

neuropsychological studies described above, the engagement of MTL and striatum may differ owing to procedural details. However, it is tempting to speculate that the late-learning deficit observed in patients with MTL lesions may be due to the absence of processes occurring early in training in normal individuals: that is, the MTL may not be strictly necessary for early learning, but its contributions during early learning may facilitate later learning. Evidence that MTL lesions may change the nature of early learning on classification tasks comes from Reber *et al.*²⁰, who found that MTL-lesioned patients showed normal early learning on the weather-prediction task but were impaired on transfer tests that required flexible use of knowledge acquired during learning.

The importance of MTL in early learning on the weather-prediction task was predicted by computational modelling of cortico-hippocampal interaction during learning^{6–8}. These simulation studies predict that the classification task used here may require a significant degree of MTL mediation, owing to the many cue–cue configural regularities present in the stimulus set, but that the MTL is active only during early phases of learning in which new representations of the stimuli are being developed. Once new stimulus representations have been established early in learning, the model predicts that MTL involvement will decrease. These predictions are confirmed by the present fMRI results demonstrating that the MTL is initially active (during putative development of new stimulus representations) but subsequently decreases (after such representations stabilize). Analogous electrophysiological results have been found in rabbit eyeblink conditioning in which the hippocampal activity related to conditioned responses appears early in learning but disappears with extended training²¹. This computational theory interprets both the earlier animal data and the present human imaging data as implying an interaction between the hippocampus and other brain structures, in which the hippocampus has a modulatory role in learning by developing new stimulus representations during early phases of training which are used by the striatum to develop complex stimulus–response associations.

Competition between memory systems may reflect an adaptive mechanism for optimizing behaviour depending upon learning demands. Some animal studies^{3,22,23} have shown that the MTL and striatum acquire different types of information during learning: the MTL appears to acquire flexible, relational knowledge (such as spatial relationships) whereas the striatum acquires inflexible stimulus–response associations. These differences have been most directly examined using a cross-maze learning task in rats. The rat is first taught to run to a particular arm of the maze from a given starting point, and is then tested by starting from a different location. The MTL supports place learning (in which the rat runs towards the previously rewarded spatial location from the new starting position) whereas the striatum supports response learning (in which the rat makes the previously rewarded motor response from the new starting position). Normal rats exhibit a transition from early reliance upon MTL-based place learning strategies to later reliance upon striatal-based response learning strategies^{22,23}. Pharmacologic facilitation of MTL or suppression of striatum increases the later engagement of place strategies, whereas facilitation of striatum or suppression of MTL increases the early engagement of response strategies^{22,23}, suggesting that these structures are locked in a ‘zero-sum’ competition to control behaviour.

The interpretation of apparent task-related deactivations is complicated by the fact that such deactivations can reflect either active decreases during the task of interest or increases during the baseline task. In the present case there are two findings that suggest that the deactivations do not reflect increases during the baseline activity. First, MTL deactivation during the classification task has been found using three very different baseline tasks: counting the number of cards presented⁵, simple button press (experiment 1), and visual fixation (experiment 2). Second, activation was modu-

lated depending on task demands in experiment 1 (FB versus PA tasks) when the baseline task remained constant across these two tasks. Thus, we believe that the results reflect a specific decrease in fMRI signal during the classification task. However, the neurophysiological basis of this decrease remains unknown. In particular, there is current debate as to whether both excitatory and inhibitory synaptic activity result in increased fMRI signal^{24,25}. Neural modelling suggests that the relation between fMRI signal and inhibition may depend upon features of local neural circuitry, such as the degree of recurrence²⁶. Our results clearly demonstrate that MTL deactivation during FB learning reflects a response to task demands, but further work is necessary to understand fully the synaptic correlates of these signals.

Thus we have shown that learning involves competition between the MTL-based and striatal-based memory systems in the human brain. These findings provide a direct link to convergent findings of competitive memory systems in animal learning, and suggest that competition may serve as a mechanism to arbitrate between two fundamentally incompatible requirements of learning: the need for flexibly accessible knowledge (supported by the MTL) and the need to learn fast, automatic responses in specific situations (supported by the striatum). □

Methods

Subjects

A total of 40 right-handed subjects participated in the studies described here: 26 in experiment 1 (6 males, age range: 18–31) and 14 in experiment 2 (2 males, age range 19–33). All gave informed consent according to procedures approved by Massachusetts General Hospital.

Task and stimuli

In experiment 1, subjects participated in four scans lasting 468 s, during which they alternated every 26 s between the classification and baseline tasks. Subjects in the feedback-based group were asked to perform the classification task on each trial by pressing one of two buttons, and received feedback (category label) on each trial after 4 s. Those in the paired-associate group simply pressed a button to denote stimulus onset, and received the category label at trial onset (see Fig. 1). During the baseline task in both conditions, subjects were presented with a set of stimuli along with the instruction ‘PRESS’ and were asked simply to press a button to denote that the stimulus had appeared. For subjects in the PA group, learning was tested immediately after completion of the last scan by presenting all of the card combinations twice and asking subjects to classify each; no feedback was given.

In experiment 2, individual trials were presented as in experiment 1 (FB condition) but the interstimulus interval was varied across trials; the sequence of interstimulus intervals was determined by optimizing the design matrix for estimation of the fMRI response²⁷. Event-related response was estimated in comparison to a baseline of visual fixation. Two scans lasting 450 s were performed with this task, with 48 trials occurring during each scan. Behavioural data were collected using a button box in the scanner; data for 3 subjects in experiment 1 (2 in the FB group and 1 in the PA group) were lost owing to computer malfunction.

Stimuli in both experiments were a set of four cards with geometric shapes (after ref. 10), any combination of which could appear on a given trial (except for all cards or none). Category labels were probabilistically associated with stimuli, with cue strengths of the individual cards specified as 80%, 60%, 40% and 20% respectively. These probabilities are slightly more extreme than those used in previous studies of this task^{5,10}, in order to encourage more consistent learning. Accuracy was determined according to a probability-maximizing metric, that is, responses that matched the most correct response for a given set of cards were counted as correct.

MRI acquisition

Twenty-one axial slices (5 mm thick, 1 mm gap) were collected at 1.5 T (Siemens Sonata; experiment 1) or 3 T (Siemens Allegra; experiment 2) using a gradient-echo echo-planar pulse sequence (repetition time, 2,000 ms; echo time, 40 ms (at 1.5 T) and 30 ms (at 3 T), field of view, 200 mm; matrix size, 64 × 64). High-resolution T1-weighted scans (MP-RAGE; Siemens) were acquired for anatomical localization.

Data analysis

Preprocessing and statistical analysis of the data were performed using SPM99 software (Wellcome Dept of Cognitive Neurology). Preprocessing included slice timing correction (experiment 2 only), motion correction, normalization to the MN1305 stereotactic space (interpolating to 3 mm cubic voxels) and spatial smoothing with an 8-mm gaussian kernel. Statistical analysis was performed using the general linear model. Global signal scaling was not applied, in order to prevent spurious deactivations. The design in experiment 1 was modelled using a boxcar function convolved with a canonical

haemodynamic response function, whereas the event-related design in experiment 2 was modelled using a canonical haemodynamic response and its temporal derivative²⁸. Comparisons of interest were implemented as linear contrasts. This analysis was performed individually for each subject, and contrast images for each subject were used in a second-level analysis treating subjects as a random effect. Statistical maps were thresholded at $P < 0.005$ uncorrected for multiple comparisons with an extent threshold of 5 voxels. Tests on hypothesized regions were corrected for multiple comparisons ($P < 0.05$) using the gaussian-field small-volume correction in SPM99 after thresholding the statistical map at $P < 0.025$ with an extent threshold of 5 voxels. Parametric analysis was performed by including an additional regressor that modelled an exponentially changing HRF with a time constant of 100 s (ref. 17).

For functional connectivity analysis, signal from a seed region in the MTL (centred at $-33, -30, -18$, with a 6-voxel radius) was compared with all other voxels across subjects using simple correlation analysis, with a statistical parametric map determined for regions of significant positive or negative correlation²⁹.

Received 2 July; accepted 24 October 2001.

- Gabrieli, J. D. Cognitive neuroscience of human memory. *Annu. Rev. Psychol.* **49**, 87–115 (1998).
- Eichenbaum, H. E. & Cohen, N. J. *From Conditioning to Conscious Recollection: Memory Systems of the Brain* (Oxford Univ. Press, New York, 2001).
- Eichenbaum, H., Fagan, A., Mathews, P. & Cohen, N. J. Hippocampal system dysfunction and odor discrimination learning in rats: Impairment or facilitation depending on representational demands. *Behav. Neurosci.* **102**, 331–339 (1988).
- Packard, M. G., Hirsh, R. & White, N. M. Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: evidence for multiple memory systems. *J. Neurosci.* **9**, 1465–1472 (1989).
- Poldrack, R. A., Prabakaran, V., Seger, C. & Gabrieli, J. D. E. Striatal activation during cognitive skill learning. *Neuropsychology* **13**, 564–574 (1999).
- Gluck, M. & Myers, C. Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus* **3**, 491–516 (1993).
- Gluck, M. A., Oliver, L. M. & Myers, C. E. Late-training amnesic deficits in probabilistic category learning: a neurocomputational analysis. *Learn. Mem.* **3**, 326–340 (1996).
- Gluck, M. A. & Myers, C. E. *Gateway to Memory: An Introduction to Neural Network Modeling of the Hippocampus and Learning* (MIT Press, Cambridge, Massachusetts, 2001).
- Gluck, M. A. & Bower, G. H. From conditioning to category learning: an adaptive network model. *J. Exp. Psychol. Gen.* **117**, 227–247 (1988).
- Knowlton, B., Squire, L. R. & Gluck, M. A. Probabilistic classification in amnesia. *Learn. Mem.* **1**, 106–120 (1994).
- Mishkin, M., Malamut, B. & Bachevalier, J. in *Neuropsychology of Memory* (eds Squire, L. R. & Butters, N.) 287–296 (Guilford, New York, 1984).
- Squire, L. R. Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* **99**, 195–231 (1992).
- Knowlton, B. J., Mangels, J. A. & Squire, L. R. A neostriatal habit learning system in humans. *Science* **273**, 1399–1402 (1996).
- Knowlton, B. J. *et al.* Dissociations within nondeclarative memory in Huntington's disease. *Neuropsychology* **10**, 538–548 (1996).
- Winocur, G. & Weiskrantz, L. An investigation of paired-associate learning in amnesic patients. *Neuropsychologia* **14**, 97–110 (1976).
- Warrington, E. K. & Weiskrantz, L. Amnesia: A disconnection syndrome? *Neuropsychologia* **20**, 233–248 (1982).
- Buchel, C., Holmes, A. P., Rees, G. & Friston, K. J. Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *Neuroimage* **8**, 140–148 (1998).
- Kim, J. J. & Baxter, M. G. Multiple brain-memory systems: the whole does not equal the sum of its parts. *Trends Neurosci.* **24**, 324–330 (2001).
- Hopkins, R., Myers, C., Shohamy, D. & Gluck, M. Impaired category learning in hypoxic subjects with hippocampal damage. *Soc. Neurosci. Abstr.* **27**, 347.5 (2001).
- Reber, P. J., Knowlton, B. J. & Squire, L. R. Dissociable properties of memory systems: differences in the flexibility of declarative and nondeclarative knowledge. *Behav. Neurosci.* **110**, 861–871 (1996).
- Sears, L. L. & Steinmetz, J. E. Acquisition of classically conditioned-related activity in the hippocampus is affected by lesions of the cerebellar interpositus nucleus. *Behav. Neurosci.* **104**, 681–692 (1990).
- Packard, M. G. & McGaugh, J. L. Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiol. Learn. Mem.* **65**, 65–72 (1996).
- Packard, M. G. Glutamate infused posttraining into the hippocampus or caudate-putamen differentially strengthens place and response learning. *Proc. Natl Acad. Sci. USA* **96**, 12881–12886 (1999).
- Waldvogel, D. *et al.* The relative metabolic demand of inhibition and excitation. *Nature* **406**, 995–998 (2000).
- Magistretti, P. J. & Pellerin, L. Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. *Phil. Trans. R. Soc. Lond. B* **354**, 1155–1163 (1999).
- Tagamets, M. & Horwitz, B. Interpreting PET and fMRI measures of functional neural activity: the effects of synaptic inhibition on cortical activation in human imaging studies. *Brain Res. Bull.* **54**, 267–273 (2001).
- Dale, A. M. Optimal experimental design for event-related fMRI. *Hum. Brain Map.* **8**, 109–114 (1999).
- Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N. & Dale, A. M. Stochastic designs in event-related fMRI. *Neuroimage* **10**, 607–619 (1999).
- Horowitz, B., McIntosh, A. R., Haxby, J. V. & Grady, C. L. Network analysis of brain cognitive function using metabolic and blood flow data. *Behav. Brain Res.* **66**, 187–193 (1995).

Acknowledgements

We thank N. J. Cohen, R. Raizada and A. Wagner for helpful comments. This work was supported by the Alafi Family Foundation and the Athinoula A. Martinos Center for Biomedical Imaging.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.P. (e-mail: poldrack@nmr.mgh.harvard.edu).

Inhibitory PAS domain protein is a negative regulator of hypoxia-inducible gene expression

Yuichi Makino^{*†}, Renhai Cao[‡], Kristian Svensson^{§||}, Göran Bertilsson^{§||}, Mikael Åsman[¶], Hiroto Tanaka[†], Yihai Cao[‡], Anders Berkenstam^{§||} & Lorenz Poellinger^{*}

^{*} Department of Cell and Molecular Biology, Medical Nobel Institute;

[‡] Department of Microbiology and Tumor Biology; and [¶] Center for Genomics and Bioinformatics, Karolinska Institutet, S-171 77 Stockholm, Sweden

[†] Division of Clinical Immunology, Advanced Clinical Research Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

[§] Pharmacia Corporation, S-112 87 Stockholm, Sweden

Alteration of gene expression is a crucial component of adaptive responses to hypoxia. These responses are mediated by hypoxia-inducible transcription factors (HIFs)^{1,2}. Here we describe an inhibitory PAS (Per/Arnt/Sim) domain protein, IPAS, which is a basic helix-loop-helix (bHLH)/PAS protein structurally related to HIFs. IPAS contains no endogenous transactivation function but demonstrates dominant negative regulation of HIF-mediated control of gene expression. Ectopic expression of IPAS in hepatoma cells selectively impairs induction of genes involved in adaptation to a hypoxic environment, notably the vascular endothelial growth factor (VEGF) gene, and results in retarded tumour growth and tumour vascular density *in vivo*. In mice, IPAS was predominantly expressed in Purkinje cells of the cerebellum and in corneal epithelium of the eye. Expression of IPAS in the cornea correlates with low levels of expression of the VEGF gene under hypoxic conditions. Application of an IPAS antisense oligonucleotide to the mouse cornea induced angiogenesis under normal oxygen conditions, and demonstrated hypoxia-dependent induction of VEGF gene expression in hypoxic corneal cells. These results indicate a previously unknown mechanism for negative regulation of angiogenesis and maintenance of an avascular phenotype.

When screening mouse expressed sequence tag (EST) databases using hidden Markov model profiles³ for HIF homology search, we identified an EST clone encoding a putative new protein, designated IPAS, containing a bHLH PAS motif. DNA sequence analysis demonstrated that IPAS complementary DNA contains an open reading frame of 921 nucleotides, encoding a polypeptide of 307 amino acids (Fig. 1a). Alignment analysis of this amino-acid sequence with known bHLH/PAS factors showed high similarity to HIF-1 α ⁴ and HLF (HIF- α -like factor, also known as endothelial PAS domain protein 1, EPAS-1)^{5,6} in the amino-terminal bHLH domain (75% and 76% identity, respectively; Fig. 1b), and to a lesser extent within the PAS region (34% and 36% in the PAS A domain, and 40% and 36% in the PAS B domain, respectively; Fig. 1b).

^{||} Present addresses: Karobia AB, S-141 57 Huddinge, Sweden (A.B.); Neuronova AB, S-114 33 Stockholm, Sweden (G.B.); and Leica Microsystems AB, S-191 27 Sollentuna, Sweden (K.S.)

Notably, IPAS lacks the sequence corresponding to the carboxy-terminal region of HIF-1 α and HLF, in which two transactivation domains (N-TAD and C-TAD) have been identified (Fig. 1b).

Northern blot analysis of poly(A)⁺ RNA from a variety of mouse tissues demonstrated that IPAS was expressed predominantly in the eye and at lower levels in the cerebellum and the cerebrum, indicating a tissue-restricted expression pattern of IPAS messenger RNA (Fig. 1c). To characterize the spatial expression pattern of IPAS in the eye and cerebellum, *in situ* hybridization assays were performed. Intense IPAS expression was observed in the epithelial cell layer of the cornea (Fig. 2a, b) and with lower intensity in the layers of ganglion cells, inner nuclear cells, and rods and cones of the retina (Fig. 2e, f). Expression of HIF-1 α mRNA was detected at low levels in the epithelium of the cornea (Fig. 2c, d), demonstrating strikingly dominant expression of IPAS over HIF-1 α in these cells. HIF-1 α was also expressed in the same layers of retina where IPAS expression was observed (Fig. 2g, h). In the cerebellum, expression of IPAS was limited to the Purkinje cell layer (Fig. 2i, j), whereas HIF-1 α did not show any spatially defined expression throughout the sections (Fig. 2k, l). Finally, both IPAS and HIF-1 α mRNAs were detected as weak diffuse signals over nonspecific background levels in certain areas of the cerebrum (data not shown).

The structural similarity of IPAS to HIFs and the colocalization of IPAS and HIF-1 α in mouse cornea prompted us to investigate the potential role of IPAS in regulation of transcriptional responses to hypoxia. We performed a transient transfection assay in HeLa cells using a luciferase reporter gene driven by a hypoxia response element (HRE) in the absence or presence of transiently expressed IPAS. In the absence of IPAS, incubation of the cells under hypoxic

(1% O₂) conditions resulted in induction of reporter gene activity, reflecting the induced transactivation function of endogenous HIFs. Transient expression of IPAS reduced this hypoxia-induced activation response (Fig. 3a, b). Moreover, hypoxia-dependent activation of the reporter gene by co-expression of either HIF-1 α (Fig. 3a) or HLF (Fig. 3b) was impaired in a dose-dependent manner by IPAS. These results suggest that IPAS acts as a dominant negative regulator of HIF-1 α - and HLF-mediated gene expression. IPAS had no effect on hypoxia-induced protein stabilization of HIF-1 α or HLF⁷⁻¹⁰ (Fig. 3c), indicating that IPAS targeted regulatory steps further downstream than protein stabilization in HIF-mediated signal transduction.

To further investigate the role of IPAS in regulation of HIF function, we generated mouse hepatoma Hepa 1c1c7 cells stably expressing IPAS. Expression of IPAS mRNA was not detected in the wild-type cells (Fig. 3d). As expected^{11,12}, under hypoxic conditions these cells showed markedly increased expression of phosphoglycerate kinase-1 (PGK1) and VEGF mRNA. In response to hypoxia, IPAS-expressing Hepa (Hepa IPAS) cells showed decreased levels of hypoxia-dependent activation of these genes (45% and 48% reduction of the induction response, respectively; Fig. 3d). This IPAS-mediated effect seemed to be at the transcriptional level, because activation of a transiently transfected HRE-driven reporter gene by hypoxia was significantly lower in Hepa IPAS cells than in wild-type cells (data not shown). Reporter gene activation was even suppressed in Hepa IPAS cells after transient overexpression of HIF-1 α , indicating that IPAS impairs productive interaction between HIF-1 α and the HRE. In strong support of this model, HRE-specific DNA-binding activity by the HIF-1 α -Arnt heterodimeric complex

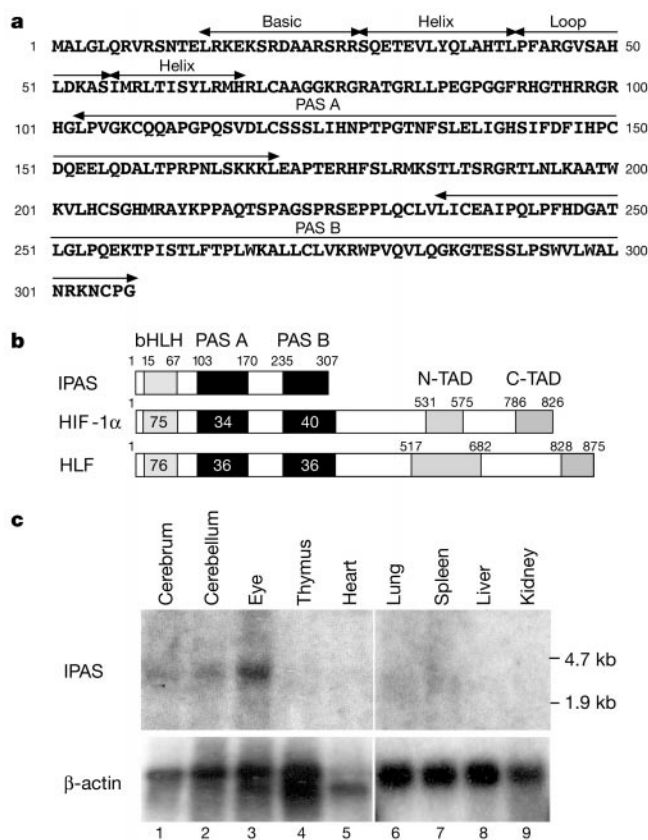


Figure 1 IPAS is a bHLH/PAS factor. **a**, Amino-acid sequence of mouse IPAS. **b**, Comparison of IPAS with mouse HIF-1 α (AAC53461) and mouse HLF (BAA20130), showing per cent amino-acid identity within the bHLH and PAS domains. **c**, RNA blot analysis of IPAS expression in adult mouse tissues. The positions of RNA markers are shown on the right in kilobases.

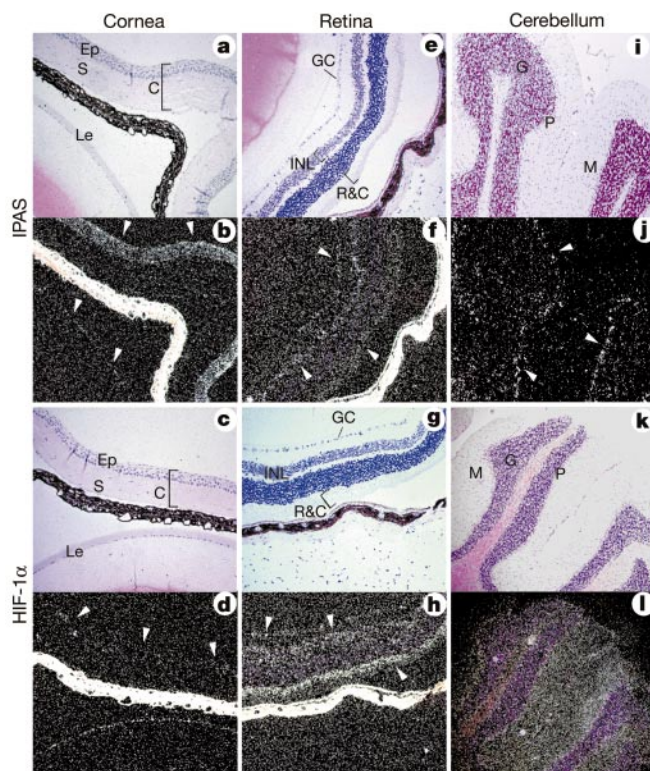


Figure 2 *In situ* hybridization analysis of IPAS expression in mouse cornea, retina and cerebellum. Sections of cornea (**a–d**), retina (**e–h**) and cerebellum (**i–l**) of adult mouse were hybridized with antisense RNA probes of mouse IPAS (**a, b, e, f, i and j**) or HIF-1 α (**c, d, g, h, k and l**). Light-field (**a, c, e, g, i and k**) and dark-field (**b, d, f, h, j and l**) views are shown. C, cornea; Ep, epithelium; S, substantia propria; LE, lens epithelium; GC, granular cell layer; INL, inner nuclear layer; R&C, rods and cones; G, granular layer; P, Purkinje cells; M, molecular layer.

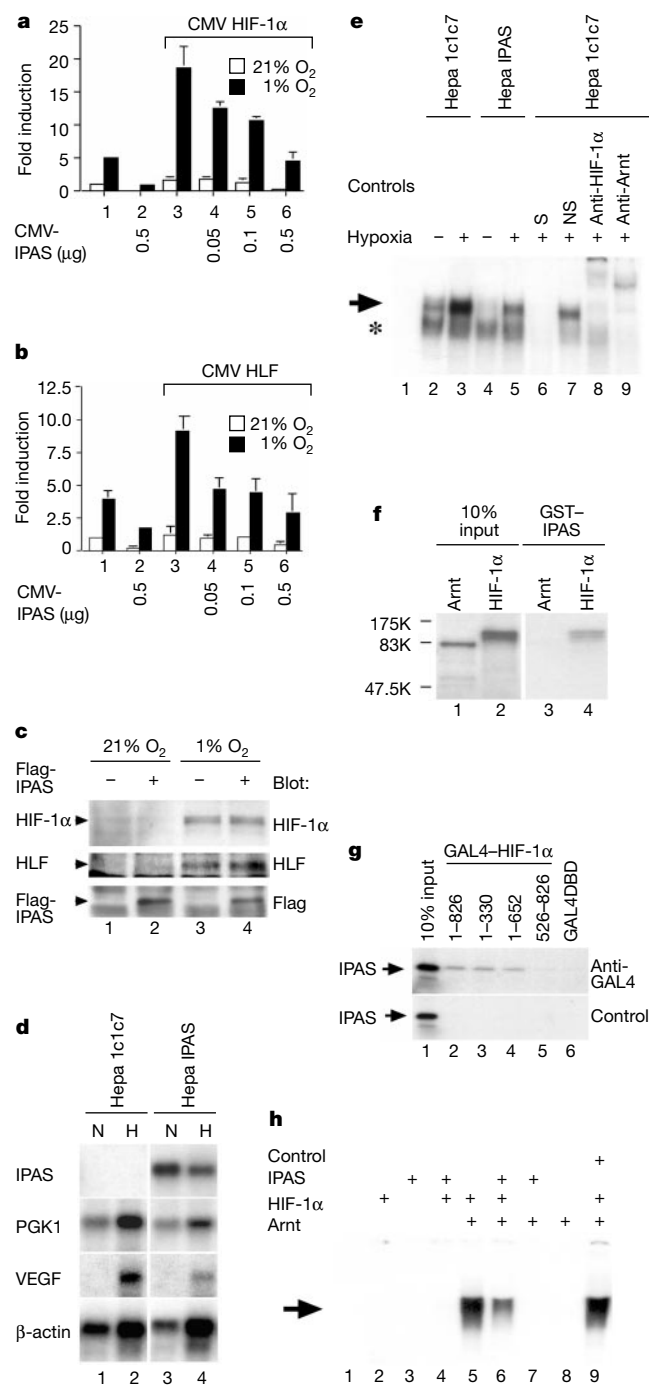


Figure 3 IPAS is a dominant negative regulator of HIFs. **a, b**, IPAS inhibits reporter gene activation mediated by HIF-1 α (**a**) or HLF (**b**) in hypoxic HeLa cells (means $\pm$ s.d. of fold induction relative to cells transfected with reporter gene alone). **c**, Stability of HIF-1 α and HLF proteins is not affected by IPAS in HeLa cells. **d**, Hypoxia-inducible gene expression is impaired in cells overexpressing IPAS. RNA was prepared from cells incubated at normoxia (N) or hypoxia (H). **e**, IPAS inhibits the DNA-binding activity of HIF-1 α . Asterisk, constitutive HRE-binding activity. Arrow, HIF-1 α Arnt DNA complex¹⁶. Controls are competition experiments with unlabelled HRE (S) or unrelated sequence (NS), or supershift with indicated antibodies. **f**, GST-IPAS interacts physically with *in vitro* translated HIF-1 α , as assessed by immunoprecipitation with anti-GST antibodies. The positions of protein markers are shown on the left in relative molecular mass (M). **g**, The N terminus of HIF-1 α mediates heterodimerization with IPAS. Co-immunoprecipitation of ³⁵S-labelled IPAS with GAL4-fused HIF-1 α fragments by anti-GAL4 antibodies. GAL4DBD, Gal4 DNA-binding domain. **h**, The IPAS-HIF-1 α heterodimer fails to bind the HRE motif. Combinations of *in vitro* translated IPAS, HIF-1 α , Arnt and control lysate were analysed.

was lower in nuclear extracts from either normoxic (normal oxygen concentration, 21% O₂) or hypoxic IPAS-expressing cells than in corresponding nuclear extracts from wild-type cells (Fig. 3e). The dioxin (aryl hydrocarbon) receptor that mediates gene regulation in response to xenobiotic chemicals is also a member of the bHLH/PAS transcription factor family and shares the dimerization partner factor Arnt with HIF-1 α ¹³. In contrast to hypoxia-induced gene expression, dioxin-induced expression of the dioxin receptor target gene for cytochrome P-4501A1 was not perturbed in the Hepa IPAS cells (data not shown). Taken together, IPAS seems to preferentially target HIFs to act as a dominant negative regulator of hypoxia-induced gene expression.

We next examined whether the inhibitory action of IPAS was mediated by direct interaction with HIF-1 α or Arnt. In co-immunoprecipitation assays, a glutathione S-transferase (GST)-IPAS fusion protein was coprecipitated with HIF-1 α but not with Arnt, demonstrating specific physical interaction between IPAS and HIF-1 α (Fig. 3f). The immunoprecipitation assays further demonstrated that N-terminal structures of HIF-1 α (amino acids 1–330), mainly composed of the bHLH/PAS motif, mediated association with IPAS (Fig. 3g). In support of this model, mammalian two-hybrid assays using as a bait a GAL4 fusion protein spanning the bHLH/PAS domain of HIF-1 α demonstrated interaction with an IPAS-VP16 fusion protein. Consistent with the *in vitro* precipitation experiments, GAL4-IPAS failed to show any interaction with Arnt-VP16 *in vivo* (data not shown). To characterize the functional consequences of the IPAS-HIF-1 α interaction, we performed DNA-binding experiments. In agreement with the reduced levels of HIF-1 α -dependent DNA-binding activity observed in nuclear extracts of cells stably

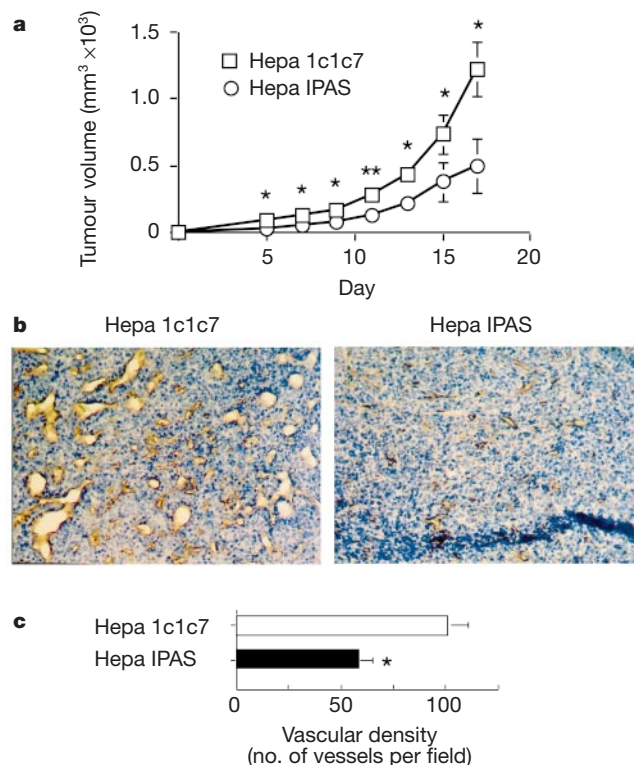


Figure 4 Expression of IPAS in hepatoma cells retards tumour growth. **a**, Tumour volumes (mean $\pm$ s.d.) after subcutaneous implantation of wild-type Hepa 1c1c7 or Hepa IPAS cells into immunodeficient SCID mice. Asterisk ($P < 0.001$) and double asterisk ($P < 0.01$) indicate significant differences between the cell lines. **b, c**, Vascular density. At day 18 after implantation, tumour tissues were immunohistochemically stained for expression of the CD31 antigen (**b**), and stained vessels were counted under a light microscope in at least six random fields (**c**). Asterisk ($P < 0.001$) indicates a significant difference from wild-type cell-derived tumours.

expressing IPAS (Fig. 3e), the IPAS–HIF-1 α complex did not bind to an HRE probe, and the HRE-binding activity of the HIF-1 α –Arnt heterodimer was impaired when exposed to IPAS (Fig. 3h). These results demonstrate that IPAS forms an abortive complex with HIF-1 α with regard to its ability to recognize the HRE motif.

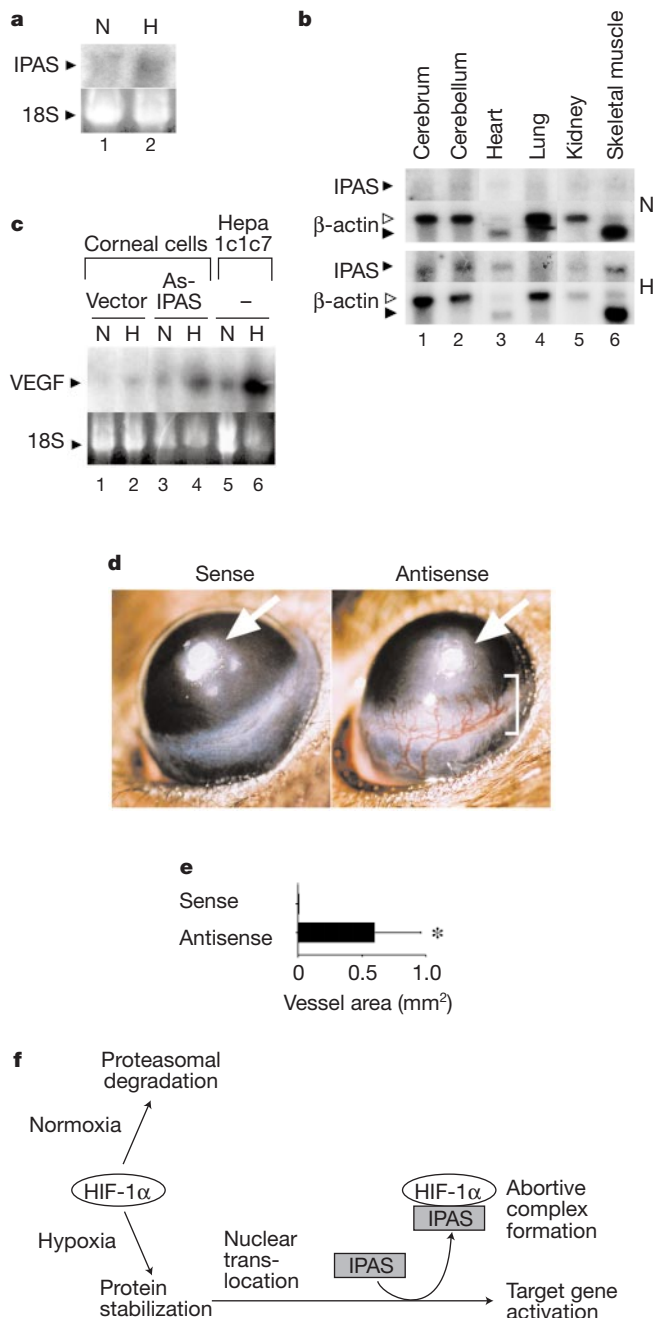


Figure 5 Evidence that IPAS functions as an inhibitor of angiogenesis in the cornea. **a**, Primary cultures of mouse corneal epithelium cells were incubated under normoxic (N) or hypoxic (H) conditions for 24 h, and total RNA was analysed for IPAS expression by RNA blotting. **b**, Hypoxia-inducible expression of IPAS mRNA in various mouse tissues. Open and filled arrowheads, 2.0-kb and 1.8-kb forms of β -actin mRNA, respectively. **c**, VEGF mRNA expression in hepatoma or primary corneal cells transfected with empty vector or an antisense (AS) IPAS vector. **d**, Ocular angiogenesis studies after implantation of polymers containing IPAS antisense or sense oligonucleotides into micropockets of the cornea (arrow). Corneal neovascularization is indicated by the bracket. **e**, Area of neovascularization, presented as mean $\pm$ s.d. Asterisk, $P < 0.001$ compared with the group treated with sense oligonucleotide. **f**, Model of negative regulation of HIF-1 α function by IPAS.

Given the dominant negative effect of IPAS on HIF-1 α function, we examined a potential role of IPAS in negative regulation of angiogenesis by determining both the growth rates and vascular density of tumours formed after subcutaneous implantation of either wild-type Hepa 1c1c7 or Hepa IPAS cells in immuno-incompetent SCID mice. IPAS-expressing hepatoma cells produced tumours that showed a slower growth rate than that of the wild-type cells ($P < 0.001$ – 0.01 at all times of observation; Fig. 4a). In agreement with these results, the tumours derived from the IPAS-expressing cells had a significantly ($P < 0.001$) lower vascular density than the wild-type tumours, as assessed by staining tumour samples with CD31 antibodies (Fig. 4b, c).

We wondered about the biological significance of the striking expression of IPAS in corneal epithelium. A characteristic hallmark of the cornea is its total avascularity, creating the transparency required for clear vision. After overnight closure of the eye during sleep, the corneal environment is known to generate conditions of hypoxia^{14,15} that in other cells are sufficient to stimulate HIF-1 α -dependent gene expression. However, gene activation and neovascularization do not occur under these conditions in the cornea, and the underlying mechanism of negative regulation is unknown. We therefore investigated whether IPAS negatively regulates VEGF expression in the cornea. We analysed expression of VEGF mRNA in primary cultures of mouse corneal epithelium cells. In agreement with the *in situ* hybridization experiments, we observed low but detectable levels of IPAS mRNA under normoxic conditions. Interestingly, exposure of the cells to hypoxia for 24 h resulted in enhanced levels of IPAS mRNA expression (Fig. 5a). To assess the potential regulation of IPAS expression by hypoxia in other tissues, we exposed mice to hypoxia (6% O₂) for 6 h. Consistent with our earlier observations (Fig. 1c), low levels of IPAS expression were observed in most tissues of control mice kept at normoxia. However, IPAS mRNA expression was induced by hypoxia not only in the corneal epithelium but clearly (2.2- to 3.9-fold induction) also in the cerebrum, cerebellum, heart and skeletal muscle (Fig. 5b). These results indicate that, under conditions of hypoxia or ischaemia, IPAS may modulate regulatory responses of hypoxia-inducible genes in these tissues also.

In sharp contrast to wild-type mouse Hepa 1c1c7 cells¹⁶, primary corneal cells transfected with an empty expression vector demonstrated almost undetectable basal levels of VEGF mRNA, and hypoxia produced only a very modest induction response (Fig. 5c). To test whether IPAS expression determines this low degree of hypoxia responsiveness, primary corneal cells were transiently transfected with an antisense IPAS expression vector to downregulate IPAS expression levels. In contrast to the cells transfected with an empty control vector, the introduction of the antisense IPAS vector resulted in elevation of basal levels of VEGF mRNA expression and significant hypoxia-inducible expression of the VEGF gene (Fig. 5c). These results strongly suggest that IPAS is important in the negative regulation of both basal expression level and hypoxia-dependent activation of the VEGF gene in corneal epithelium.

To investigate whether IPAS is involved in the regulation of angiogenesis in the cornea, scrambled, sense or antisense IPAS oligonucleotides were adsorbed onto slow-release pellets and implanted into mouse corneas. The stimulation of blood vessel growth from the limbus region was recorded after 5 days without any reduction of atmospheric oxygen levels. Notably, this analysis revealed significantly ($P < 0.001$) induced neovascularization in the corneas of mice treated with the IPAS antisense oligonucleotide (Fig. 5d, e). Thus, these results suggest that the physiological conditions of hypoxia that occur in the cornea during, for example, sleep^{14,15} may induce vascular growth when IPAS expression is inhibited.

Because dysregulation or enhanced activation of HIF-1 α function has been implicated in a variety of pathological conditions includ-

ing induced tumour growth rates^{9,17,18} and inflammatory angiogenesis¹⁹, it may be physiologically relevant for certain cells or tissues to be able to functionally counteract the activated form of HIF-1 α . Our data suggest that the hypoxia-inducible IPAS may provide such a cell-type-specific or conditional negative regulatory strategy by the formation in the nucleus of a nonfunctional complex with HIF-1 α (Fig. 5f). This mechanism may be relevant for regulation of HIF function not only in the cornea but also in certain other tissues exposed to hypoxia. In agreement with this model, IPAS shows predominantly nuclear localization under both normoxic and hypoxic conditions, as assessed by assays with green fluorescent fusion protein (data not shown). The mode of negative regulation of HIF-1 α signalling by IPAS resembles the mechanism of repression of the function of MyoD and related bHLH differentiation factors by Id²⁰. In summary, our data indicate that negative regulation of HIF-1 α function by IPAS defines an anti-angiogenic mechanism in hypoxic cells, as illustrated by the profound inhibitory effect of IPAS on VEGF gene expression and angiogenesis in the cornea. □

Methods

Database search and isolation of IPAS cDNA

Hidden Markov model profiles were built with the HMMER 1.8.3 software²¹ from nucleotide sequences corresponding to the PAS domain of a selected number of bHLH/PAS factors. This model was used to search EST databases, and one of the identified ESTs (GenBank accession number AA028416) was obtained, sequenced and designated IPAS. pcDNA3 IPAS was constructed by insertion of an *EcoRI*–*NotI* fragment from pT7T3D IPAS (GenBank AA028416) into pcDNA3 (Invitrogen), and this vector was used to subclone IPAS into the pCMV Flag vector²², pGEX-4T-3 (Amersham Pharmacia), pCMX GAL4 or pCMX VP-16, to create expression vectors for GST, GAL4 and VP16 fusion proteins. To construct an antisense IPAS expression plasmid, full-length IPAS cDNA was inserted in the inverted direction into pcDNA3.

Cell culture and transfection experiments

Hepa 1c1c7, HeLa and COS7 cells were obtained from the American Type Culture Collection (ATCC). Hepa IPAS cells were established by stable transfection of Hepa 1c1c7 cells with pEFIREspuro IPAS and puromycin (5 μ g ml⁻¹) selection. In assays with luciferase reporter genes, 0.5 μ g of reporter plasmid and indicated amounts of expression plasmids were used for transfection using cationic liposomes (Invitrogen). Conditions for treatment with hypoxia or 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) of the cells have been described previously¹⁶. Protein expression was monitored in whole-cell extracts by immunoblot assays with HIF-1 α (Novus), HLF (Novus) and Flag epitope (Sigma) antibodies essentially as described previously^{8,22}. Mouse primary corneal epithelial cells were isolated from 6-week-old C57Bl6/J healthy mice killed by a lethal dose of CO₂. The eyes were enucleated and the corneal tissue was dissected in DMEM medium supplemented with 10% bovine calf serum under a stereomicroscope. Tissue masses were placed onto a gelatin-coated tissue-culture plate and incubated in the same medium supplemented with 3 ng ml⁻¹ human recombinant fibroblast growth factor-2 (FGF-2). After incubation for 8 days, corneal epithelial cells were treated with trypsin, and single-cell suspensions were used for subsequent assays.

RNA blot and *in situ* hybridization analysis

Poly(A)⁺ RNA samples (4.5 μ g) from various tissues of C57Bl6 mice, Hepa 1c1c7 or Hepa IPAS cells were obtained by the guanidiumthiocyanate method followed by oligo(dT) purification (Dyna) and analysed by RNA blotting using ³²P-labelled cDNA probes of mouse IPAS (nucleotides 623–897), PGK1 (426–771), VEGF3 (24–466), CYP1A1 (874–1,199) and β -actin (930–1,075). *In situ* hybridization of tissue sections from 8-week-old C57Bl6 mice using ³⁵S-labelled mouse IPAS or HIF-1 α antisense RNA probes was performed as previously described²³.

In vitro protein interaction assays

Nuclear extracts from normoxic or hypoxic cells were prepared and analysed by electrophoretic mobility shift assays as described before¹⁶. GST–IPAS or GAL4 fusion proteins spanning various fragments of HIF-1 α were generated by translation in the presence or absence of ³⁵S-labelled methionine in rabbit reticulocyte lysate (Promega). Equal concentrations of ³⁵S-labelled, *in vitro* translated Arnt, HIF-1 α and IPAS were incubated with GST–IPAS or GAL4–HIF-1 α proteins for 1 h at room temperature, then incubated with anti-GST (Amersham Pharmacia) or anti-GAL4 (Upstate) antibodies coupled to protein A-Sepharose (Amersham Pharmacia) for another hour at room temperature. After brief centrifugation, co-immunoprecipitated proteins were analysed by SDS–polyacrylamide gel electrophoresis and autoradiography.

Corneal micropocket assay and hypoxic exposure

The mouse corneal angiogenesis assay was performed as described²⁴. Micropellets of sucrose aluminium sulphate and hydron polymers containing 2.5 μ g of phosphorothio-

ate-coupled IPAS antisense (5′-TCACGCGCTGCAGCCCCAACGCCAT-3′), sense (5′-ATGGCGTTGGGGCTGCAGCGCGTGA-3′) or scrambled oligonucleotides were implanted into corneal pockets of 6-week-old male C57Bl6/J mice. Six mice and twelve corneas were used in each group. The vascular area of the corneas of all the mice was measured by slit-lamp stereomicroscopy on day 5 after pellet implantation, and statistically evaluated by Student's two-tailed *t*-test. In hypoxia experiments, 8-week-old mice were either kept at normoxia or exposed to normobaric hypoxia (6% O₂) for 6 h in an airtight chamber flushed with a mixture of nitrogen and room air. All animal studies were reviewed and approved by the animal care and use committee of the Stockholm Animal Board.

Tumour studies in mice

Male 8–10-week old immuno-incompetent SCID mice were used for tumour implantation. About 1 $\times$ 10⁶ wild-type Hepa 1c1c7 or stably IPAS-expressing hepatoma cells were implanted subcutaneously in each mouse. Six and seven mice were used in the treated and control groups, respectively. Primary tumours were measured with digital calipers on the days indicated. Tumour volumes were calculated according to the formula width² $\times$ length $\times$ 0.52, as previously reported²⁴. At day 18 after implantation, tumour tissues were cut out and fixed with 4% formalin in phosphate-buffered saline for 24 h. Tissues were imbedded in paraffin and stained with a rat anti-mouse antibody against CD31 antigen²⁴; positive signals were counted under a light microscope in at least six random fields at $\times$ 40 magnification.

Received 16 May; accepted 4 October 2001.

- Bunn, H. F. & Poyton, R. O. Oxygen sensing and molecular adaptation to hypoxia. *Physiol. Rev.* **76**, 839–885 (1996).
- Wenger, R. H. Mammalian oxygen sensing, signalling and gene regulation. *J. Exp. Biol.* **203**, 1253–1263 (2000).
- Krogh, A., Brown, M., Mian, I. S., Sjolander, K. & Haussler, D. Hidden Markov models in computational biology. Applications to protein modeling. *J. Mol. Biol.* **235**, 1501–1531 (1994).
- Wang, G. L., Jiang, B. H., Rue, E. A. & Semenza, G. L. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O₂ tension. *Proc. Natl Acad. Sci. USA* **92**, 5510–5514 (1995).
- Ena, M. *et al.* A novel bHLH-PAS factor with close sequence similarity to hypoxia-inducible factor 1 α regulates the VEGF expression and is potentially involved in lung and vascular development. *Proc. Natl Acad. Sci. USA* **94**, 4273–4278 (1997).
- Tian, H., McKnight, S. L. & Russell, D. W. Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. *Genes Dev.* **11**, 72–82 (1997).
- Semenza, G. L. HIF-1 and human disease: one highly involved factor. *Genes Dev.* **14**, 1983–1991 (2000).
- Kallio, P. J., Wilson, W. J., O'Brien, S., Makino, Y. & Poellinger, L. Regulation of the hypoxia-inducible transcription factor 1 α by the ubiquitin-proteasome pathway. *J. Biol. Chem.* **274**, 6519–6525 (1999).
- Maxwell, P. H. *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271–275 (1999).
- Wiesener, M. S. *et al.* Induction of endothelial PAS domain protein-1 by hypoxia: characterization and comparison with hypoxia-inducible factor-1 α . *Blood* **92**, 2260–2268 (1998).
- Li, H., Ko, H. P. & Whitlock, J. P. Induction of phosphoglycerate kinase 1 gene expression by hypoxia. Roles of Arnt and HIF1 α . *J. Biol. Chem.* **271**, 21262–21267 (1996).
- Forsythe, J. A. *et al.* Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1. *Mol. Cell Biol.* **16**, 4604–4613 (1996).
- Poellinger, L. in *Inducible Gene Expression* (ed. Baeuerle, P. A.) 177–205 (Birkhäuser, Boston, 1995).
- Thakur, A., Willcox, M. D. & Stapleton, F. The proinflammatory cytokines and arachidonic acid metabolites in human overnight tears: homeostatic mechanisms. *J. Clin. Immunol.* **18**, 61–70 (1998).
- Sack, R. A., Beaton, A. R. & Sathe, S. Diurnal variations in angiotensin in human tear fluid: a possible role in prevention of corneal neovascularization. *Curr. Eye Res.* **18**, 186–193 (1999).
- Gradin, K. *et al.* Functional interference between hypoxia and dioxin signal transduction pathways: competition for recruitment of the Arnt transcription factor. *Mol. Cell Biol.* **16**, 5221–5231 (1996).
- Ravi, R. *et al.* Regulation of tumor angiogenesis by p53-induced degradation of hypoxia-inducible factor 1 α . *Genes Dev.* **14**, 34–44 (2000).
- Zundel, W. *et al.* Loss of PTEN facilitates HIF-1-mediated gene expression. *Genes Dev.* **14**, 391–396 (2000).
- Li, J. *et al.* PR 39, a peptide regulator of angiogenesis. *Nature Med.* **6**, 49–55 (2000).
- Norton, J. D., Deed, R. W., Craggs, G. & Sablitzky, F. Id helix-loop-helix protein in cell growth and differentiation. *Trends Cell Biol.* **8**, 58–65 (1998).
- Eddy, S. R. Multiple alignment using hidden Markov models. *Ismb* **3**, 114–120 (1995).
- Tanimoto, K., Makino, Y., Pereira, T. & Poellinger, L. Mechanism of regulation of the hypoxia-inducible factor-1 α by the von Hippel-Lindau tumor suppressor protein. *EMBO J.* **19**, 4298–4309 (2000).
- Bertilsson, G. *et al.* Identification of a human nuclear receptor defines a new signaling pathway for CYP3A induction. *Proc. Natl Acad. Sci. USA* **95**, 12208–12213 (1998).
- Cao, R. *et al.* Suppression of angiogenesis by a novel inhibitor, K1-5, generated by plasmin-mediated proteolysis. *Proc. Natl Acad. Sci. USA* **96**, 5728–5733 (1999).

Acknowledgements

We thank Y. Fujii-Kuriyama and I. Pongratz for reagents. This work was supported by grants from the Swedish Medical Research Council, Pharmacia Corporation, Swedish Cancer Society, the Human Frontiers Science Program, the Uehara Memorial Foundation, and the Japanese Ministry of Education, Culture, Sports, Science and Technology.

Correspondence and requests for materials should be addressed to L.P. (e-mail: lorenz.poellinger@cmb.ki.se). The cDNA sequence of IPAS is deposited in GenBank under accession number AF416641.

Extensive surface diversity of a commensal microorganism by multiple DNA inversions

Corinna M. Krinos*, Michael J. Coyne*, Katja G. Weinacht*, Arthur O. Tzianabos*, Dennis L. Kasper**† & Laurie E. Comstock*

* Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

† Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

The dynamic interactions between a host and its intestinal microflora that lead to commensalism are unclear. Bacteria that colonize the intestinal tract do so despite the development of a specific immune response by the host¹. The mechanisms used by commensal organisms to circumvent this immune response have yet to be established. Here we demonstrate that the human colonic microorganism, *Bacteroides fragilis*, is able to modulate its surface antigenicity by producing at least eight distinct capsular polysaccharides—a number greater than any previously reported for a bacterium—and is able to regulate their expression in an on–off manner by the reversible inversion of DNA segments containing the promoters for their expression. This means of generating surface diversity allows the organism to exhibit a wide array of distinct surface polysaccharide combinations, and may

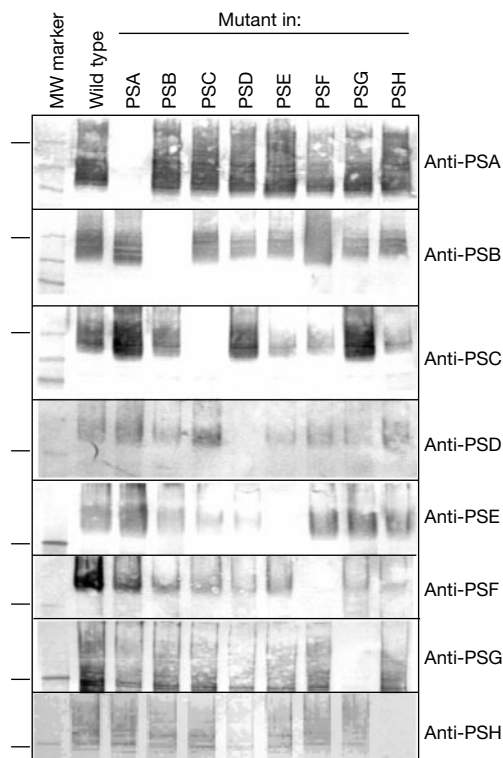


Figure 1 Western immunoblot analysis of reactivity of monospecific antisera with wild type and mutants. Bacterial lysates were added to a 4–20% gradient SDS-polyacrylamide gel, transferred to polyvinylidene fluoride, and probed with monospecific antiserum equivalent to a 1/3,000 dilution of the original antiserum. The reactivity of each of the eight capsule mutants with each of the eight specific antisera demonstrate the specificity of each antiserum for a single polysaccharide. The horizontal line indicated for each molecular weight (MW) standard identifies the standard relative molecular mass of 200,000 (M_r , 200K).

have broad implications for how the predominant human colonic microorganisms, the *Bacteroides* species, maintain an ecological niche in the intestinal tract.

The capsular polysaccharide complex of *B. fragilis* was previously shown to be composed of at least three distinct capsular polysaccharides: PSA, PSB and PSC^{2–4}. A fourth polysaccharide biosynthesis locus was located by transposon mutagenesis (GenBank accession number AF366394), which we designate the PSD locus. These four loci each contain two open reading frames (ORFs) just upstream of the polysaccharide biosynthesis genes designated *upxY* and *upxZ* (where x is a , b , c or d , depending on the locus), which demonstrate 38–78% similarity for the *upxY* products and 44–72% similarity for the *upxZ* products. Our data suggest that the *upxY* products transcriptionally regulate their respective polysaccharide biosynthesis locus (C.M.K., D. Lipsett and L.E.C., unpublished work). The function of the *upxZ* products in the regulation of

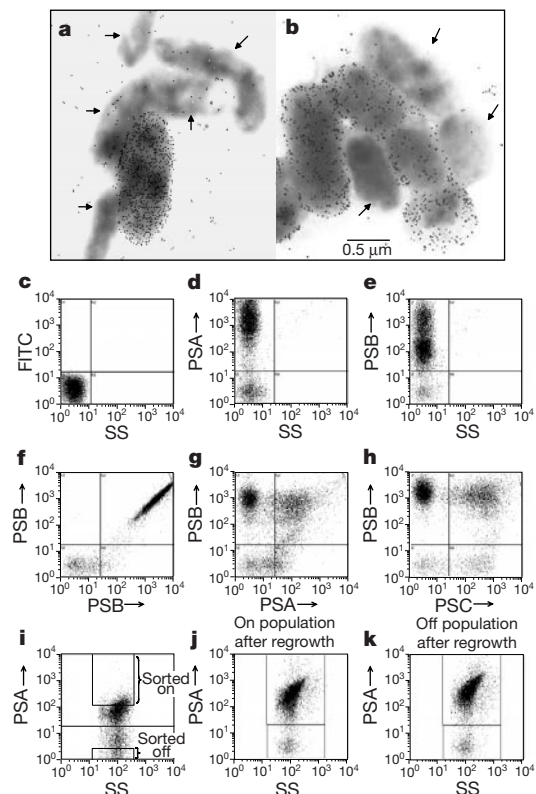


Figure 2 Phase-variable expression of capsular polysaccharides of *B. fragilis*.

a, b, Electron micrographs showing the on or off phenotype of PSA (**a**), detected with monoclonal antibody (mAb) CE3, or PSB (**b**), using mAb G9, followed by incubation with protein G labelled with colloidal gold. Bacteria that are not labelled with these polysaccharide-specific mAbs are highlighted with arrows. **c–k**, Flow cytometric analyses demonstrating polysaccharide phase variation. Secondary antibodies were an FITC-labelled anti-mouse immunoglobulin- γ (IgG; y axis) and a phycoerythrin-labelled anti-rabbit IgG (x axis), where applicable. SS, side scatter. **c**, Irrelevant antibody control with wild-type NCTC9343. **d**, Expression of PSA as monitored with mAb CE3. **e**, Expression of PSB as monitored with mAb G9. Double labelling experiments are shown in panels **f–h**. **f**, Control demonstrating PSB-specific mAb in combination with the PSB-specific rabbit antiserum showing almost all bacteria either unlabelled or doubly labelled. **g**, Coexpression of PSB (mAb G9; y axis) with PSA (rabbit polyclonal; x axis). **h**, Coexpression of PSB (mAb G9; y axis) with PSC (rabbit polyclonal; x axis). **i**, Bacteria expressing high levels of PSA were sorted from those that were negative for PSA expression as indicated (reanalysis of sorted populations demonstrated greater than 92% purity for the positive population and 97% purity for the negative population). **j**, Bacteria sorted as PSA positive (from **i**) were able to switch to PSA negative after regrowth. **k**, Bacteria sorted as PSA negative (from **i**) were able to switch to PSA positive after regrowth.

their downstream loci have yet to be elucidated. Sequencing of the *B. fragilis* NCTC9343 chromosome by the Sanger Centre's microbial pathogen group is near completion. To determine if *B. fragilis* synthesizes additional capsular polysaccharides, we searched the sequence data (provided at http://www.sanger.ac.uk/Projects/B_fragilis/) for the presence of additional paralogues of *upxY* and *upxZ*, and retrieved four new adjacent gene pairs (designated *upeY/upeZ* through to *uphY/uphZ*). The extent of similarity of the eight UpxY or UpxZ gene products is shown in Supplementary Information Fig. 1. Our analysis of the DNA regions downstream of each of these four new gene pairs revealed closely linked genes whose products are similar to proteins involved in polysaccharide biosynthesis.

To determine whether these four loci encode products involved in the synthesis of capsular polysaccharides, we performed a mutational analysis. A polar mutation was created downstream of the *upxZ* gene of each locus by insertion of a nonreplicating plasmid by homologous recombination. Monospecific antisera were prepared by adsorption of antiserum to wild-type *B. fragilis* NCTC9343 with each mutant as described². Each of these four mutants is unable to synthesize a surface molecule of high molecular mass, with a heterogeneous size distribution characteristic of the capsular poly-

saccharides of *B. fragilis* (Fig. 1, bottom four panels). These four antibody preparations and monospecific antisera to PSA, PSB, PSC and PSD²⁻⁴ were tested for reactivity with each of the eight capsular polysaccharide mutants of *B. fragilis*. Each of the eight antibody preparations is specific for a single polysaccharide, as each of these antibody preparations fails to react with only one of the eight mutants (Fig. 1). Each of these four loci contains genes encoding homologues of an initiating glycosyltransferase, a polysaccharide flippase and a polymerase, providing additional evidence that these regions synthesize distinct polysaccharides—designated PSE, PSF, PSG and PSH. The genes with homology to polysaccharide biosynthesis genes in these four new loci (PSE: 16 genes, 15,775 base pairs (bp); PSF: 18 genes, 15,256 bp; PSG: 21 genes, 22,732 bp; and PSH: 16 genes, 14,730 bp) combined with the genes from the four previously characterized regions, totals a significant amount of DNA in a bacterial genome (133 kb) dedicated to the synthesis of capsular polysaccharides.

Electron microscopy and flow cytometry revealed that each of the eight capsule-specific antibody preparations labelled only a portion of the bacteria from any given culture, whereas some bacteria were not labelled at all (Fig. 2a–e; see also Supplementary Information Fig. 2). This result was also obtained when the culture was derived

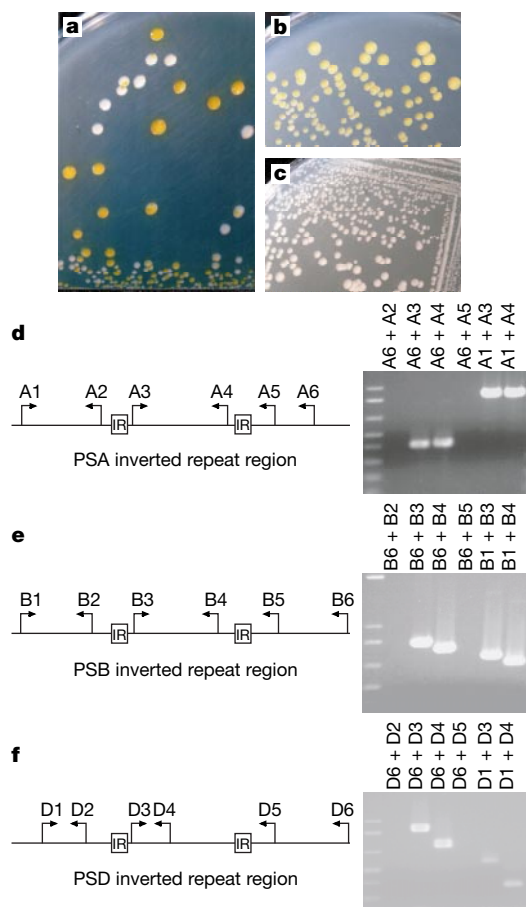


Figure 3 DNA inversions between inverted repeats account for phase-variable promoter activity. **a**, A 420-bp region upstream of *upaY* confers variable expression of XylE as indicated by yellow and white colonies. **b**, **c**, A 217-bp region containing DNA internal to the outermost ends of each inverted repeat of the PSA invertible region is unable to phase vary and is constitutively on (**b**) or off (**c**) when cloned in each orientation in relation to *xylE*. **d**–**f**, PCR analysis of wild-type NCTC9343 chromosome showing that the DNA between the inverted repeats (IR) of the PSA (**d**), PSB (**e**) and PSD (**f**) regions invert. Arrows indicate primer orientation.

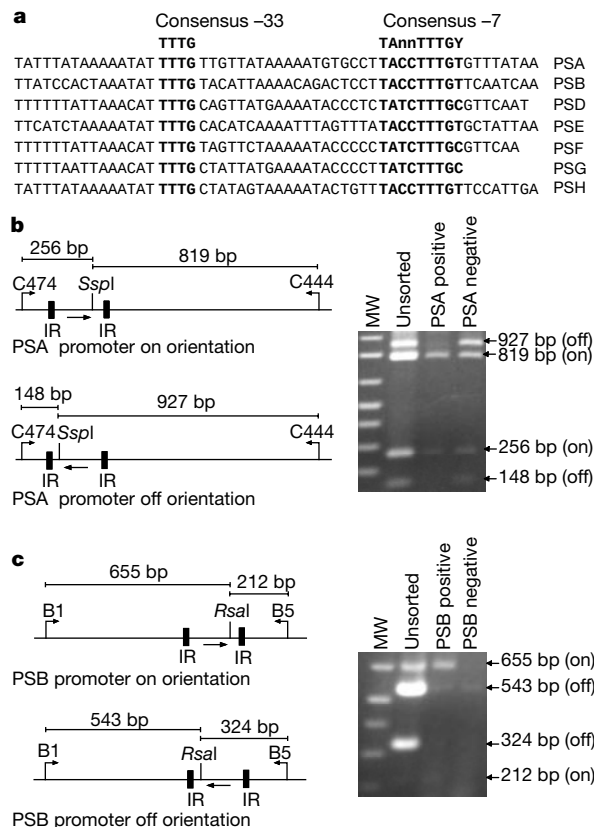


Figure 4 Polysaccharide expression correlates with the orientation of the invertible promoter regions. **a**, Alignment of the promoter regions of all seven invertible regions in the on orientation with the *B. fragilis* consensus promoter sequence. The downstream inverted repeat of each region begins immediately following the last base reported here. **b**, **c**, PCR/digestion technique demonstrates PSA and PSB promoter orientations from various populations (see Methods). IR, inverted repeat. **b**, Bacteria expressing PSA preferentially have the promoter upstream of the PSA locus in the on orientation, and those that are not expressing PSA demonstrate a greater percentage in the promoter off orientation. **c**, Most of the bacteria from the initial unsorted culture (12.4% expressing PSB by FACS analysis) have the PSB promoter oriented off (543-bp band predominates over the 655-bp band). Similarly, those bacteria sorted as PSB positive yield a denser band at 655 bp, and those bacteria sorted as PSB negative yield a denser band at 543 bp.

from a single colony, suggesting that the variable expression of the polysaccharides is attributable to regulation rather than mutation. The percentages of bacteria positive for expression of any individual polysaccharide varied greatly between experiments from 10% to 95% of the population.

Double-labelling experiments were performed using PSA-, PSB- and PSC-specific antibodies. When bacteria were analysed for concomitant expression of PSB and either PSA or PSC, four distinct populations were observed: positive only for the first polysaccharide; positive only for the second polysaccharide; negative for both polysaccharides; and positive for both polysaccharides (Fig. 2g, h). Similar results were obtained when the PSA- or PSC-specific monoclonal antibodies were used with heterogeneous antisera (data not shown). For the polysaccharide combinations analysed, expression of each polysaccharide seems to be regulated independently. Therefore, this organism may theoretically display up to 256 distinct polysaccharide combinations.

Fluorescence-activated cell sorting (FACS) was used to determine whether capsule expression shifts between the on and off states. Bacteria representing the top 25% of a PSA-on population and the bottom 25% of a PSA-off population were collected, regrown to mid-log phase, and reanalysed by flow cytometry for PSA expression (Fig. 2i). An originally PSA-on or PSA-off population yielded mixed populations with both PSA-on and PSA-off phenotypes (Fig. 2j, k). The same results were obtained with PSB and PSC on and off populations (data not shown), demonstrating that these polysaccharides undergo phase variation defined as a reversible on-off phenotype. Switching between on to off and off to on occurs at similar frequencies of at least 0.01 per cell per generation, but these switching frequencies seem to be variable and are probably subject to environmental regulation.

Our previous analysis of the PSC biosynthesis locus of strain 638R revealed a single promoter 240-bp upstream of *upcY* (C.M.K., D. Lipsett and L.E.C., unpublished work). We predicted that the promoters of the other seven polysaccharide loci would also be located upstream of their respective *upxY* paralogue. A 420-bp region upstream of *upaY* was analysed using a reporter plasmid that uses the enzymatic activity encoded by *xylE* to detect *B. fragilis* promoters by converting a colony from white to yellow when substrate is applied to the plate. The resulting *B. fragilis* transconjugants revealed both XylE positive (yellow) and XylE negative (white) colonies (Fig. 3a). When yellow colonies were re-streaked and analysed, the resulting colonies were a mixture of both yellow and white; as was the case when a white colony was re-streaked. Analysis of this 420-bp region revealed 19-bp inverted repeats separated by 193 bp of intervening DNA. Six of the seven other polysaccharide loci also contain very similar inverted repeats separated by a comparable amount of DNA upstream of their *upxY* counterpart (Supplementary Information Table 1). We proposed that the DNA between each set of inverted repeats reversibly inverts to position a promoter in the correct or incorrect orientation for transcription of the downstream genes. Polymerase chain reaction (PCR) analysis of the PSA, PSB and PSD inverted repeat regions showed that a primer between the inverted repeats yields a PCR product when paired with either a forward primer upstream of the inverted repeats or a reverse primer downstream of the inverted repeats. However, DNA outside the inverted repeats was shown not to invert (Fig. 3). A similar analysis of the regions upstream of *upeY*, *upfY*, *upgY* and *uphY* confirmed that the DNA between each of these inverted repeats also inverts (Supplementary Information Fig. 3). Of the eight polysaccharide biosynthesis loci, only the PSC locus contains no inverted repeats upstream of its *upxY* homologue, and only for the PSC locus were we unable to demonstrate inversion of its promoter region. Our data, however, suggest that phase variation of PSC is also regulated at the transcriptional level (C.M.K. and L.E.C., unpublished work).

To determine exactly where the inversion of the PSA region

occurs, the plasmid containing the 420-bp region was isolated from *B. fragilis*, and a clone with the invertible region in the opposite orientation was sequenced, revealing that the DNA inverts exactly between the inverted repeats. Our subsequent analysis of the sequence data from the Sanger Centre identified runs in which the DNA between the inverted repeats of the PSB, PSD, PSE and PSH regions is present in each orientation, allowing us to determine that these invertible regions also invert exactly between the inverted repeats. The PSF and PSG invertible regions are present in the database in only one orientation. We amplified, by PCR, and sequenced these regions in the opposite orientation and confirmed that they also invert exactly between their respective inverted repeats.

The data presented above demonstrate that the DNA between the inverted repeats of the PSA locus contains a functional promoter when oriented in one direction in relation to *xylE*. To determine which orientation confers promoter activity, DNA between and including all but 7 bp of each PSA inverted repeat was cloned into the reporter vector in both orientations in relation to *xylE*. PCR analyses confirmed our hypothesis that clones that do not contain the entire inverted repeat elements do not invert in *B. fragilis*. Each clone exhibited a single yellow or white phenotype (Fig. 3b, c). Similar constructs were created for each of the other six invertible regions and analysis of the resulting clones in *B. fragilis* demonstrated that each provided promoter activity in only one orientation. The levels of XylE enzymatic activity from the yellow clone of each of the seven regions was at least 100-fold greater than the XylE activity from the corresponding, oppositely oriented white clone. These analyses demonstrate that the invertible regions upstream of the PSA, PSB, PSD, PSE, PSF, PSG and PSH biosynthesis loci contain functional promoters and demonstrate the direction in which these regions must be oriented for transcription of the downstream polysaccharide biosynthesis genes.

Analysis of the sequences of the seven invertible regions in the on orientation demonstrates that each aligns perfectly with a recently reported consensus promoter sequence for *B. fragilis*⁵ (Fig. 4a). To determine if the promoter on orientation correlates with the phenotypic expression of the polysaccharide encoded by the downstream genes, we used a PCR/digestion technique modelled after one used to quantify the orientation of the *fim* switch region of *Escherichia coli*⁶. We used FACS to separate bacteria expressing or not expressing PSA. Chromosomal DNA of these two populations was subjected to PCR/digestion as shown in Fig. 4b. Before sorting, 78.8% of the bacteria from the culture were PSA positive, as determined by FACS analysis. PCR/digestion of this unsorted population showed that a greater percentage of these bacteria had the PSA promoter oriented on. Of the bacteria sorted as PSA positive, PCR/digestion revealed a much denser band at 819 bp (promoter on orientation) compared with the 927-bp band. The inverse was true for those bacteria that sorted as PSA negative. The results were similar for the orientation of the putative PSB promoter and the bacteria that were sorted as PSB positive or PSB negative (Fig. 4c). Therefore, on orientations of the PSA and PSB promoter directly correlate with the expression of these polysaccharides.

Very few studies have been directed towards understanding the mechanisms that commensal organisms use to persist in the changing environment of the human intestine. This study reports the expression by *B. fragilis* of an extensive and unprecedented array of distinct capsular polysaccharides; a process that is subject to phase variation dictated by invertible promoter regions upstream of the polysaccharide biosynthesis loci. Preliminary data indicate that these inversions are mediated by site-specific recombinases. Phase variation of type 1 fimbriae of *E. coli* occurs by a similar mechanism^{7,8}; however, DNA inversions have not been shown to regulate phase variation of other bacterial polysaccharides. The variable expression of this multitude of polysaccharides is probably

paramount for *B. fragilis* to maintain a long-term commensal relationship in the human colon. As *B. fragilis* is the anaerobic species most frequently isolated from clinical infections, and the capsular polysaccharides are instrumental in the disease process, phase variation may also contribute to the pathogenic potential of this organism. □

Methods

Construction of insertion mutants

Sequences of primers used in this study are listed in Supplementary Information Table 2. DNA used for homologous recombination for each of the four mutants was amplified by PCR from NCTC9343 using the following primers: PSE: U1-F, UpE-R; PSF: U2-F, UpF-R; PSG: UY5-F, UpG-R; PSH: Y6F2, UpH-R. These 2.1–3.3-kb products were digested with *Bam*HI and cloned into pJST55 (ref. 9). Plasmids were introduced into *B. fragilis* NCTC9343 by mobilization from *E. coli* and cointegrates were selected using erythromycin.

Demonstration of inversion and PCR/digestion technique

Primers used to demonstrate inversion—in the order of upstream primer, primer between the inverted repeats and downstream primer—are as follows: PSE: UpE-1, UpE-2, UpE-3; PSF: UpF-1, UpF-2, UpF-3; PSG: UpG-1, UpG-2, UpG-3; PSH: UpH-1, UpH-2, UpH-3.

For the PCR/digestion technique in Fig. 4b, c, a PCR is performed, using chromosomal DNA from a phenotypically selected population, which amplifies the PSA or PSB invertible region and some flanking DNA. The PCR product is digested with a restriction enzyme that cleaves asymmetrically between the inverted repeat elements. When the promoter is in one orientation, the two fragments resulting from the digested PCR product differ in size from those fragments that result when the promoter is in the opposite orientation.

PSF and PSG inversion junctions

Inversion junctions were determined by sequencing PCR products resulting from amplification using the following primers that only amplify DNA in one orientation: PSF: UpF-2, UpF-3; PSG: UpG-2, UpG-3.

Creation of *xylE* reporter constructs

The 420-bp region upstream of *upaY* was amplified by PCR using primers PMA-1 and PMA-3, and was cloned into the *Bam*HI site of *xylE* reporter plasmid pLEC23. The following primer pairs were used for PCR amplification of the invertible regions lacking 7–9 bp of each inverted repeat: PSA: IRA-F, IRA-R; PSB: IRB-F, IRB-R; PSD: IRD-F, IRD-R; PSE: IRE-F, IRE-R; PSF: IRF-F, IRF-R; PSG: IRG-F, IRG-R; PSH: IRH-F, IRH-R. Each PCR product was cloned into pLEC23 in both orientations. Xyle assay was performed as described (C.M.K., D. Lipsett and L.E.C., unpublished work).

Received 6 August; accepted 3 October 2001.

- Macpherson, A. et al. A primitive T cell-independent mechanism of intestinal mucosal IgA responses to commensal bacteria. *Science* **288**, 2222–2226 (2000).
- Comstock, L. E. et al. Analysis of a capsular polysaccharide biosynthesis locus of *Bacteroides fragilis*. *Infect. Immun.* **67**, 3525–3532 (1999).
- Coyne, M. J., Kalka-Moll, W., Tzianabos, A. O., Kasper, D. L. & Comstock, L. E. *Bacteroides fragilis* NCTC9343 produces at least three distinct capsular polysaccharides: cloning, characterization, and reassignment of the PS B and PS C biosynthesis loci. *Infect. Immun.* **68**, 6176–6181 (2000).
- Coyne, M., Tzianabos, A., Mallory, B., Kasper, D. & Comstock, L. A polysaccharide biosynthesis locus required for virulence of *Bacteroides fragilis*. *Infect. Immun.* **69**, 4342–4350 (2001).
- Bayley, D., Rocha, E. & Smith, C. Analysis of *cepA* and other *Bacteroides fragilis* genes reveals a unique promoter structure. *FEMS Microbiol. Lett.* **193**, 149–154 (2000).
- Kim, J. et al. *In vivo* phase variation of *Escherichia coli* type 1 fimbrial genes in women with urinary tract infection. *Infect. Immun.* **66**, 3303–3310 (1998).
- Abraham, J., Freitag, C., Clements, J. & Eisenstein, B. An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **82**, 5724–5727 (1985).
- Olsen, P. & Klemm, P. Localization of promoters in the *fim* gene cluster and the effect of H-NS on the transcription of *fimB* and *fimE*. *FEMS Microbiol. Lett.* **116**, 95–100 (1994).
- Thompson, J. S. & Malamy, M. H. Sequencing the gene for an imipenem-cefoxitin-hydrolyzing enzyme (CfA) from *Bacteroides fragilis*. *J. Bacteriol.* **172**, 2584–2593 (1990).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank W. Kalka-Moll for electron microscope photography and J. Daley and S. Lazo-Kallanian for assistance with FACS analysis. This work was supported by grants from the National Institutes of Health.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.E.C. (e-mail: lcomstock@channing.harvard.edu).

Barttin is a Cl[−] channel β-subunit crucial for renal Cl[−] reabsorption and inner ear K⁺ secretion

Raúl Estévez*, Thomas Boettger*, Valentin Stein*, Ralf Birkenhäger†, Edgar Otto†, Friedhelm Hildebrandt† & Thomas J. Jentsch*

* Zentrum für Molekulare Neurobiologie (ZMNH), Universität Hamburg, Falkenried 94, D-20246 Hamburg, Germany

† Universitäts-Kinderklinik, Universität Freiburg, Mathildenstrasse 1, D-79106 Freiburg, Germany

Renal salt loss in Bartter's syndrome is caused by impaired transepithelial transport in the loop of Henle. Sodium chloride is taken up apically by the combined activity of NKCC2 (Na⁺-K⁺-2Cl[−] cotransporters) and ROMK potassium channels. Chloride ions exit from the cell through basolateral ClC-Kb chloride channels. Mutations in the three corresponding genes have been identified^{1–3} that correspond to Bartter's syndrome types 1–3. The gene⁴ encoding the integral membrane protein barttin is mutated in a form of Bartter's syndrome that is associated with congenital

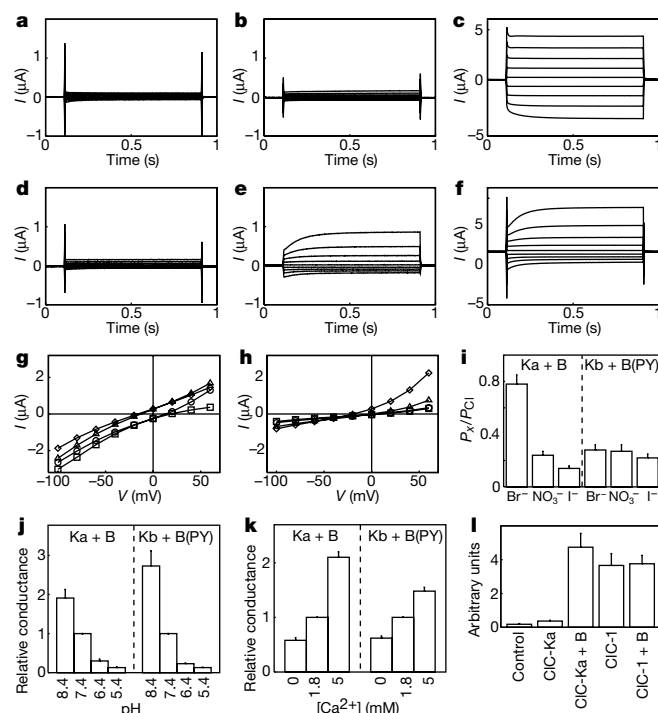


Figure 1 Functional characterization of ClC-K/barttin in *Xenopus* oocytes.

a–f, Measurements of current (*I*). Barttin (**a**), ClC-Ka (**b**) and ClC-Kb (**d**) alone gave no significant currents. ClC-Ka/barttin co-expression gave large currents (**c**), and ClC-Kb/barttin moderate currents (**e**). **f**, ClC-Kb/barttin(Y98A) currents. **g, h**, Steady-state current–voltage relationships for ClC-Ka/barttin and ClC-Kb/barttin(Y98A), respectively, in the presence of: Cl[−], diamonds; Br[−], triangles; NO₃[−], circles; I[−], squares. **i**, Permeability ratios (*P_x*/*P_{Cl}*) from reversal potentials for ClC-Ka/barttin (left) and ClC-Kb/barttin(Y98A) (right). Averages from at least two batches of oocytes, five oocytes per batch. **j**, Effect of extracellular pH on ClC-Ka/barttin (left) and ClC-Kb/barttin(Y98A). **k**, Effect of extracellular Ca²⁺ on ClC-Ka/barttin (left) and ClC-Kb/barttin(Y98A). In **j** and **k**, conductances at −20 mV were normalized to values at pH 7.4 or 1.8 mM Ca²⁺, respectively (two or three batches of oocytes, 5–9 oocytes per batch). **l**, Surface expression¹³ of epitope-tagged ClC-Ka. ClC-Ka with a cytoplasmic HA tag (left) and extracellularly tagged ClC-1 served as controls. B, barttin; B(PY), barttin(Y98A). Averages from 7–13 oocytes. Voltage (*V*) was clamped between +80 and −100 mV for 0.8 s in 20-mV steps throughout. Error bars, s.e.m.

deafness and renal failure. Here we show that barttin acts as an essential β -subunit for ClC-Ka and ClC-Kb chloride channels, with which it colocalizes in basolateral membranes of renal tubules and of potassium-secreting epithelia of the inner ear. Disease-causing mutations in either ClC-Kb or barttin compromise currents through heteromeric channels. Currents can be stimulated further by mutating a proline-tyrosine (PY) motif on barttin. This work describes the first known β -subunit for ClC chloride channels and reveals that heteromers formed by ClC-K and barttin are crucial for renal salt reabsorption and potassium recycling in the inner ear⁵.

ClC-Ka and ClC-Kb are highly homologous Cl⁻ channels that are nearly exclusively expressed in kidney⁶. *CLCNKB* mutations in Bartter's syndrome³ together with immunohistochemical results^{7,8} suggest that human ClC-Kb (the orthologue of rodent ClC-K2) mediates basolateral Cl⁻ efflux in the thick ascending limb of Henle's loop and in more distal nephron segments. Similarly, immunolocalization⁹ and the diabetes insipidus observed in *Clcnk1*^{-/-} mice¹⁰ indicate that rodent ClC-K1 (the orthologue of human ClC-Ka) is crucial for transepithelial transport in the thin ascending limb. Whereas ClC-K1 yields Cl⁻ currents on heterologous expression^{11,12}, no currents are observed with human ClC-Ka or ClC-Kb^{6,12}, suggesting that ClC-K channels may need β -subunits.

Positional cloning of the gene *BSND*, which underlies Bartter's syndrome type 4 (also named BSND; OMIM accession number 602522), identified barttin⁴. When barttin was co-expressed with ClC-Ka in *Xenopus* oocytes, large Cl⁻ currents were observed (Fig. 1c). No currents were seen with barttin, ClC-Ka or ClC-Kb alone (Fig. 1a, b, d). In oocytes, barttin/ClC-Kb co-expression gave small but detectable currents (Fig. 1e). More pronounced effects on ClC-Kb were seen in transfected tsA201 cells (see Supplementary Information A), and in oocytes injected with an activating mutant of barttin (Y98A; see below) (Fig. 1f). In oocytes, voltage activation differed between ClC-Ka/barttin and ClC-Kb/barttin (Fig. 1c, e, f). No such difference was found in mammalian cells (see Supplementary Information A). Barttin also markedly increased currents of rat ClC-K1, which yields small currents by itself^{11,12} (Fig. 2b). The effect seemed to be specific for ClC-K because currents of ClC-1, ClC-2 and ClC-5 were not changed by barttin (data not shown). Barttin

enhanced the surface expression^{13,14} of ClC-Ka, but not of the muscle channel ClC-1 (Fig. 1l).

Ion substitution revealed that both ClC-Ka/barttin and ClC-Kb/barttin were anion selective (Fig. 1g–i). As is typical for the CLC family¹⁵, they conducted Cl⁻ better than I⁻. Permeability sequences were Cl⁻ $\geq$ Br⁻ > NO₃⁻ > I⁻ for ClC-Ka/barttin, and Cl⁻ > Br⁻ = NO₃⁻ $\geq$ I⁻ for ClC-Kb/barttin (Fig. 1i). Heteromeric ClC-K/barttin channels were sensitive to extracellular pH (Fig. 1j) and Ca²⁺ (Fig. 1k). The inhibition of ClC-Ka/barttin by low extracellular pH and its stimulation by extracellular Ca²⁺ compares well to rat ClC-K1 expressed by itself^{9,12}. Compared with ClC-Ka/barttin, ClC-Kb/barttin is more sensitive to pH and less responsive to Ca²⁺.

Barttin has two hydrophobic stretches that may span the membrane (ref. 4; Fig. 2a). Epitope-tagging experiments supported this model by indicating that the first hydrophobic stretch is not a cleavable signal peptide (see Supplementary Information B). Consistent with the poor conservation between species of the second half of barttin⁴, and with the observation that all described disease-causing mutations affect the amino-terminal half⁴, large portions of the carboxy terminus could be deleted without abolishing the stimulatory effect on ClC-K1 (Fig. 2b). Truncating the protein before residue 85, however, destroyed function. The deleted segment contains a putative PY motif^{16,17} (Fig. 2a). When the critical tyrosine residue¹⁷ was mutated (Y98A), stimulation of ClC-Ka and ClC-Kb currents by barttin was enhanced (Figs 1f and 2c). Macroscopic currents did not differ qualitatively from those of wild-type heteromers. Because ClC-Kb/barttin currents are small in oocytes, this activating mutant was used to reliably investigate properties of heteromeric ClC-Kb (Fig. 1f, h–k). The effects of this mutation resemble results obtained with the Na⁺ channel ENaC^{16,17} and the Cl⁻ channel ClC-5 (ref. 14). In those cases, the increase in currents depended on an interaction of their PY motifs with WW domains of ubiquitin protein ligases^{14,18}. Future work will reveal whether WW-domain proteins are physiological regulators of ClC-K/barttin channels.

In addition to deletions and truncations, missense mutations in the N-terminal part of barttin (Fig. 2a) are associated with BSND⁴. Most of these mutations abolished or decreased the stimulatory effect on ClC-Ka or ClC-Kb (Fig. 3a, b). In the two cases that were investigated (R8L and R8W), the mutated proteins failed to increase surface expression of ClC-Ka (data not shown). Surprisingly, one missense mutation (G10S) increased ClC-Ka currents over those obtained with wild-type barttin, and did not reduce effects on ClC-Kb when studied in the framework of the activating Y98A mutant (Fig. 3a, b). Hence this mutation might cause disease by a mechanism that is not reflected in our expression system. Disease-causing missense mutations of ClC-Kb^{3,19} resulted in significant reductions or the loss of ClC-Kb/barttin currents (Fig. 3c).

In situ hybridization showed barttin expression in specific nephron segments and in the stria vascularis⁴. If barttin acts as a β -subunit for ClC-Ka and ClC-Kb, these proteins should colocalize in membranes. Immunofluorescence revealed that all nephron

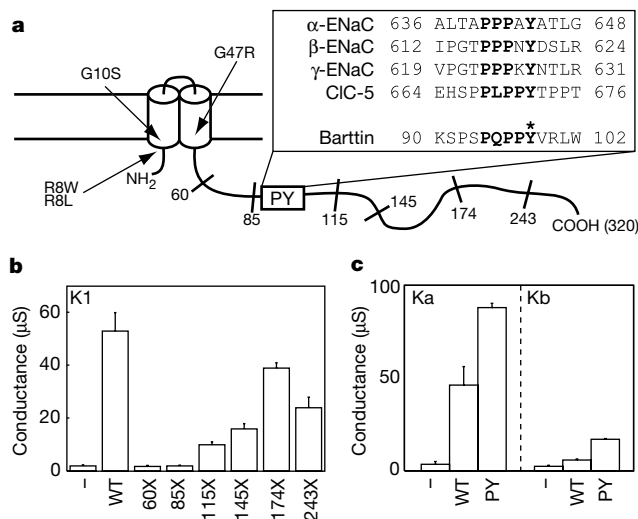


Figure 2 Basic structural and functional features of barttin. **a**, Proposed topology. Described⁴ and newly identified (G47R; data not shown) mutations are indicated. The PY motif^{16,17} (box) is compared with similar motifs of ENaC^{16,17} and ClC-5 (ref. 14) (inset). The asterisk shows the tyrosine mutated to alanine in barttin(Y98A). Truncations are indicated by slashes and the number of the stop codon. **b**, Effect of truncated barttin on ClC-K1 in oocytes. Averaged conductances at -20 mV of two batches (11 oocytes). **c**, Effect of the PY-motif mutation Y98A ('PY') on ClC-Ka (left) and ClC-Kb (right). Normalized currents of 15 oocytes from three batches. WT, wild type.

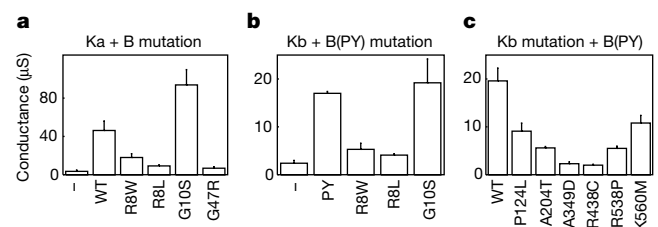


Figure 3 Functional consequences of disease-associated mutations in *Xenopus* oocytes. **a**, **b**, Effect of barttin (B) missense mutations⁴ on ClC-Ka (**a**) and ClC-Kb (**b**). **c**, Effects of *CLCNKB* missense mutations^{3,19} on ClC-Kb/barttin. In **b** and **c**, barttin(Y98A) was used to obtain sufficient currents. Conductances at -20 mV are averages of 7–25 oocytes each. WT, wild type.

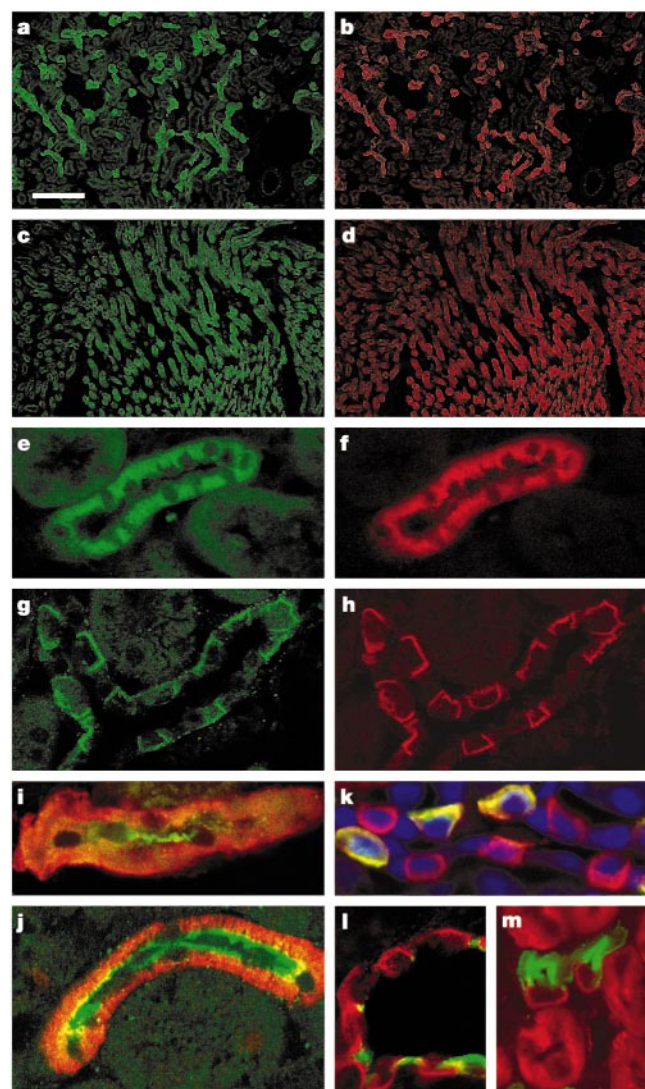


Figure 4 Barttin and CIC-K proteins in murine kidney. **a–d**, Overviews of renal cortex (**a, b**) and medulla (**c, d**) reveal strict co-expression of CIC-K (**a, c**) and barttin (**b, d**). **e, f**, Thick ascending limb (TAL) stained for CIC-K (**e**) and barttin (**f**). **g, h**, Intercalated cells of the cortical collecting duct (CCD) stained for CIC-K (**g**) and barttin (**h**). **i**, Tamm-Horsfall protein (green) identifies TAL segments that also express barttin (red). **j**, TAL showing ROMK K^+ channel (green) and barttin (red). **k**, Staining CCD for the anion exchanger AE1 (green) identifies α -intercalated cells²⁰. Barttin co-expression (red) yields a yellow stain. Cells expressing barttin and lacking AE1 are β -intercalated cells. Blue TOTO staining reveals nuclei. **l**, Aquaporin-2 (green) identifies²¹ CCD principal cells; intervening 'intercalated' cells express barttin (red). **m**, Medullary collecting ducts identified by aquaporin-2 staining (green); these and thin limbs of Henle's loop express barttin (red). Scale bar in **a** indicates 156 μ m in **a–d**, 13 μ m in **e–h** and **m**, and 10 μ m in **i–l**.

segments expressing barttin also expressed CIC-K proteins and vice versa (Fig. 4a–h). As the CIC-K antibody⁷ does not distinguish between the highly homologous CIC-K1 and CIC-K2 proteins, this suggests that barttin forms heteromers with CIC-K1 (CIC-Ka) in the thin ascending limb of Henle⁹ (Fig. 4c, d, m), and with CIC-K2 (CIC-Kb) in the thick ascending limb and more distal segments⁸ (Fig. 4e–l). Staining for barttin and CIC-K was basolateral. The Tamm-Horsfall protein (Fig. 4i) and ROMK K^+ channels (Fig. 4j) are expressed in apical membranes of the thick ascending limb, the basolateral membranes of which stained for barttin (Fig. 4f, i, j) and CIC-K (Fig. 4e). Barttin was also detected in basolateral membranes of intercalated cells of the collecting duct (Fig. 4h, k, l), which are known to express CIC-K2 (CIC-Kb) as well⁸ (Fig. 4g). On the basis of the (green) staining for the basolateral anion exchanger AE1,

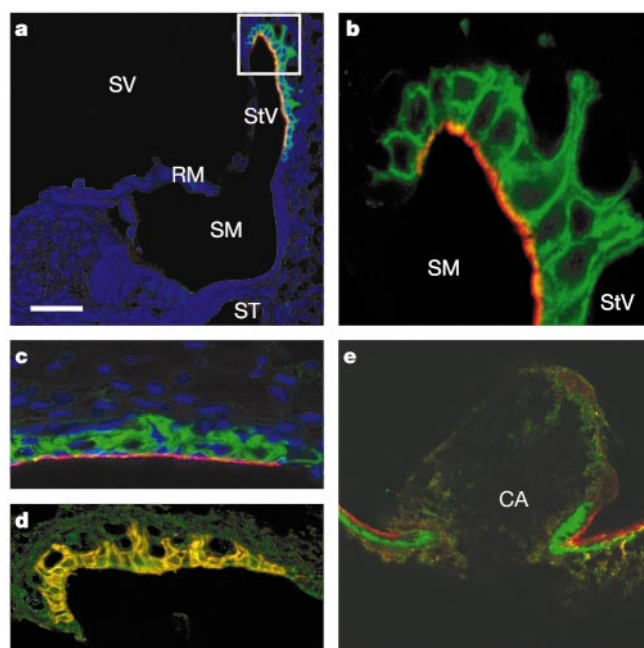


Figure 5 Barttin and CIC-K protein in the inner ear. **a**, Overview of a P0 (postnatal day 0) mouse cochlea stained for barttin (green), KCNQ1 (red) and nuclei (blue). RM, Reissner's membrane (ruptured during sectioning); SM, scala media; ST, scala tympani; StV, stria vascularis; SV, scala vestibuli. The box is enlarged in **b, c**. **b, c**, Barttin (green) and KCNQ1 (red) in adult stria vascularis. **d**, P0 cochlea co-stained for barttin (green) and CIC-K (red) reveals complete colocalization (yellow). **e**, Dark cells surrounding the crista ampullaris (CA) express barttin (green) and KCNQ1 (red). Scale bar in **a** indicates 38 μ m in **a**, 7 μ m in **b**, 22 μ m in **c** and **d**, and 31 μ m in **e**.

which identifies α -intercalated cells²⁰, both acid-secreting α -intercalated cells and base-secreting β -intercalated cells express barttin basolaterally (Fig. 4k), but intervening aquaporin-2-expressing (green) principal cells²¹ appear devoid of barttin (Fig. 4l).

In the inner ear, barttin colocalized with CIC-K in K^+ -secreting marginal cells of the stria vascularis (Fig. 5d). The basolateral staining for both proteins contrasts with the apical localization of the KCNQ1 K^+ channel (Fig. 5a–c). Previous work²² has identified CIC-K-like currents in marginal cells and correlated them with the presence of CIC-K1 messenger RNA. Our experiments with polymerase chain reaction after reverse transcription (RT-PCR) revealed that not only CIC-K1, but also CIC-K2, was expressed in the cochlea (see Supplementary Information C), indicating that both isoforms are expressed in stria vascularis. Barttin (green, Fig. 5e) was also found in K^+ -secreting vestibular dark cells, where it colocalized in basolateral membranes with CIC-K (not shown) below apical membranes that expressed KCNQ1. No balance problems were reported in BSND, but humans can adapt well to vestibular disturbances²³.

This work has identified a β -subunit for CIC-K Cl^- channels. In the kidney, CIC-K/barttin heteromers mediate Cl^- reabsorption by facilitating its basolateral efflux (Fig. 6a). In the stria, CIC-K/barttin channels drive K^+ secretion by recycling Cl^- for the basolateral NKCC1 cotransporter (Fig. 6b). This role is analogous to that of ROMK in Cl^- -reabsorbing cells of the thick ascending limb, where it recycles K^+ for the apical NKCC2 cotransporter (Fig. 6a).

Because barttin is crucial for CIC-Kb function, its inactivation results in renal salt wasting as do mutations in CIC-Kb³. However, because barttin also associates with CIC-Ka, additional symptoms are expected. These may resemble the diabetes insipidus-like phenotype observed on disrupting mouse CIC-K1 (ref. 10). Indeed, BSND patients present with more severe renal symptoms than patients having mutations in CIC-Kb^{4,24}. Unlike mutations in barttin⁴, mutations in the CIC-Kb α -subunit³ do not cause deafness, nor was deafness described in mice disrupted for CIC-K1 (ref. 10).

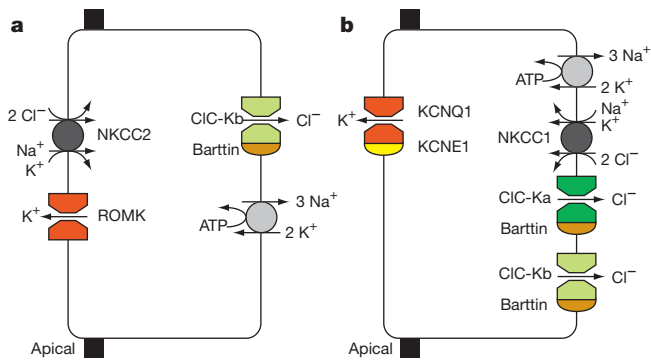


Figure 6 Model for renal Cl^- reabsorption (**a**), and for K^+ secretion in the stria vascularis (**b**). **a**, In cells of the renal thick ascending limb, apical NKCC2 cotransporters driving Cl^- uptake require ROMK to recycle K^+ . Cl^- exits through channels containing CIC-Kb α -subunits and barttin β -subunits. Mutations in all four genes cause Bartter's syndrome^{1–4}. **b**, In stria marginal cells, basolateral NKCC1 raises intracellular K^+ concentration. Parallel CIC-Ka/barttin and CIC-Kb/barttin channels recycle Cl^- . K^+ exits through channels containing KCNQ1 α -subunits and KCNE1 β -subunits. Loss of KCNQ1 (ref. 27), KCNE1 (ref. 28), NKCC1 (refs 25, 26) or barttin⁴ causes deafness. Neither loss of CIC-K1 (the orthologue of CIC-Ka)¹⁰ nor of CIC-Kb³ alone entails deafness.

We propose that both CIC-K1 and CIC-K2 are present in basolateral membranes of K^+ -secreting marginal cells (Fig. 6b). Whereas a loss of one of these α -subunits should reduce basolateral recycling of Cl^- somewhat, the mutational inactivation of the common β -subunit may abolish recycling completely. As with a disruption of the NKCC1 cotransporter^{25,26}, or with mutations in either subunit of the apical KCNQ1/KCNE1 K^+ channel^{27,28}, the resulting impairment of K^+ secretion and endolymph production probably explains the congenital deafness observed in BSND. □

Methods

Functional expression in *Xenopus* oocytes

Capped complementary RNA of CLC channels (10 ng) and barttin (5 ng) were expressed in *Xenopus* oocytes as described¹². Mutations were introduced by recombinant PCR and sequenced. Measurements were in ND96 medium (96 mM sodium chloride, 2 mM potassium chloride, 1.8 mM calcium chloride, 1 mM magnesium chloride and 5 mM HEPES buffer at pH 7.4). For anion replacement, 80 mM Cl^- was substituted by equivalent amounts of Br^- , I^- or NO_3^- . At pH 5.4 or 6.4, HEPES was replaced by 5 mM MES buffer, and at pH 8.4 by 5 mM Tris buffer.

Patch-clamp experiments of transfected cells

tsA201 cells were transiently transfected with complementary DNAs inserted into pCIneo vector with Lipofectamin (Invitrogen). Currents were measured after 2–3 days in the whole-cell mode of the patch-clamp technique using 3–5-M Ω pipettes and an Axopatch 200-A amplifier. The bath solution contained 130 mM sodium chloride, 5 mM potassium chloride, 2 mM magnesium chloride, 2 mM calcium chloride and 10 mM HEPES at pH 7.4. For ion selectivity measurements, Cl^- was exchanged for Br^- or I^- . Pipette solutions contained 130 mM caesium chloride, 5 mM sodium chloride, 2 mM calcium chloride, 2 mM magnesium chloride, 10 mM HEPES and 5 mM EGTA at pH 7.4.

Measurement of surface expression

Haemagglutinin (HA) epitopes were introduced between transmembrane domains D8 and D9 of CIC-Ka, CIC-Kb and CIC-I. This did not interfere with their ability to generate currents. Surface expression used HA antibodies and chemiluminescence was as described^{13,14}.

Antibodies

The generation of rabbit and guinea-pig antisera against barttin is described in Supplementary Information D. Published antibodies against CIC-K⁷, KCNQ1 (ref. 29), AE1 (ref. 20) and aquaporin-2 (ref. 21), and commercial antibodies against ROMK (Chemicon), Tamm-Horsfall protein (Dunn), myc epitope (9E10; American Type Culture Collection, ATCC) and HA epitope (3F10; Roche) were used. Fluorophore-coupled secondary antibodies were from Molecular Probes and Jackson.

Immunohistochemistry

Anaesthetized adult mice were perfused through the left ventricle with PBS followed by 4% paraformaldehyde (PFA) in PBS. Adult cochlea were decalcified with Rapid Bone Decalcifier (Eurobio). Inner ears from newborn mice were fixed after dissection. Tissue samples were mounted in OCT compound (Tissue Tek) for cryosections or embedded in paraffin and cut

to 8–12- μm sections. Sections were fixed in 4% PFA, 0.1% desoxycholate and 0.2% NP40 in PBS, washed, and blocked in 2% BSA, 3% goat serum and 0.5% NP40 in PBS. Antisera were applied in 2% BSA and 0.5% NP40 in PBS. Analysis was by confocal microscopy (Leica).

Received 4 October; accepted 26 October 2001.

- Simon, D. B. et al. Bartter's syndrome, hypokalaemic alkalosis with hypercalciuria, is caused by mutations in the Na-K-2Cl cotransporter NKCC2. *Nature Genet.* **13**, 183–188 (1996).
- Simon, D. B. et al. Genetic heterogeneity of Bartter's syndrome revealed by mutations in the K^+ channel, ROMK. *Nature Genet.* **14**, 152–156 (1996).
- Simon, D. B. et al. Mutations in the chloride channel gene, *CLCNKB*, cause Bartter's syndrome type III. *Nature Genet.* **17**, 171–178 (1997).
- Birkenhäger, R. et al. Mutations of *BSND* causes Bartter syndrome with sensorineural deafness and kidney failure. *Nature Genet.* **29**, 310–314 (2001).
- Jentsch, T. J. Neuronal KCNQ channel: physiology and role in disease. *Nature Rev. Neurosci.* **1**, 21–30 (2000).
- Kieferle, S., Fong, P., Bens, M., Vandewalle, A. & Jentsch, T. J. Two highly homologous members of the CIC chloride channel family in both rat and human kidney. *Proc. Natl Acad. Sci. USA* **91**, 6943–6947 (1994).
- Vandewalle, A. et al. Localization and induction by dehydration of CIC-K chloride channels in the rat kidney. *Am. J. Physiol.* **272**, F678–F688 (1997).
- Kobayashi, K., Uchida, S., Mizutani, S., Sasaki, S. & Marumo, F. Intrarenal and cellular localization of CIC-K2 protein in the mouse kidney. *J. Am. Soc. Nephrol.* **12**, 1327–1334 (2001).
- Uchida, S. et al. Localization and functional characterization of rat kidney-specific chloride channel, CIC-K1. *J. Clin. Invest.* **95**, 104–113 (1995).
- Matsumura, Y. et al. Overt nephrogenic diabetes insipidus in mice lacking the CLC-K1 chloride channel. *Nature Genet.* **21**, 95–98 (1999).
- Uchida, S. et al. Molecular cloning of a chloride channel that is regulated by dehydration and expressed predominantly in kidney medulla. *J. Biol. Chem.* **268**, 3821–3824 (1993); erratum *J. Biol. Chem.* **269**, 19192 (1994).
- Waldegger, S. & Jentsch, T. J. Functional and structural analysis of CIC-K chloride channels involved in renal disease. *J. Biol. Chem.* **275**, 24527–24533 (2000).
- Zerangue, N., Schwappach, B., Jan, Y. N. & Jan, L. Y. A new ER trafficking signal regulates the subunit stoichiometry of plasma membrane K_{ATP} channels. *Neuron* **22**, 537–548 (1999).
- Schwake, M., Friedrich, T. & Jentsch, T. J. An internalization signal in CIC-5, an endosomal Cl^- channel mutated in Dent's disease. *J. Biol. Chem.* **276**, 12049–12054 (2001).
- Jentsch, T. J., Friedrich, T., Schriever, A. & Yamada, H. The CLC chloride channel family. *Pflügers Arch. Eur. J. Physiol.* **437**, 783–795 (1999).
- Staub, O. et al. WW domains of Nedd4 bind to the proline-rich PY motifs in the epithelial Na^+ channel deleted in Liddle's syndrome. *EMBO J.* **15**, 2371–2380 (1996).
- Schild, L. et al. Identification of a PY motif in the epithelial Na channel subunits as a target sequence for mutations causing channel activation found in Liddle syndrome. *EMBO J.* **15**, 2381–2387 (1996).
- Staub, O. et al. Regulation of stability and function of the epithelial Na^+ channel (ENaC) by ubiquitination. *EMBO J.* **16**, 6325–6336 (1997).
- Konrad, M. et al. Mutations in the chloride channel gene *CLCNKB* as a cause of classic Bartter syndrome. *J. Am. Soc. Nephrol.* **11**, 1449–1459 (2000).
- Alper, S. L., Natale, J., Lodish, H. F. & Brown, D. Subtypes of intercalated cells in rat kidney collecting duct defined by antibodies against erythroid band 3 and renal vacuolar H^+ -ATPase. *Proc. Natl Acad. Sci. USA* **86**, 5429–5433 (1989).
- Nielsen, S., DiGiovanni, S. R., Christensen, E. L., Knepper, M. A. & Harris, H. W. Cellular and subcellular immunolocalization of vasopressin-regulated water channel in rat kidney. *Proc. Natl Acad. Sci. USA* **90**, 11663–11667 (1993).
- Ando, M. & Takeuchi, S. mRNA encoding 'CIC-K1, a kidney Cl^- channel' is expressed in marginal cells of the stria vascularis of rat cochlea: its possible contribution to Cl^- currents. *Neurosci. Lett.* **284**, 171–174 (2000).
- Baloh, R. W. Vertigo. *Lancet* **352**, 1841–1846 (1998).
- Jeck, N. et al. Hypokalaemic salt-losing tubulopathy with chronic renal failure and sensorineural deafness. *Pediatrics* **108**, E5 (2001).
- Delpire, E., Lu, J., England, R., Dull, C. & Thorne, T. Deafness and imbalance associated with inactivation of the secretory Na-K-2Cl co-transporter. *Nature Genet.* **22**, 192–195 (1999).
- Dixon, M. J. et al. Mutation of the Na-K-Cl co-transporter gene *Slc12a2* results in deafness in mice. *Hum. Mol. Genet.* **8**, 1579–1584 (1999).
- Neyroud, N. et al. A novel mutation in the potassium channel gene *KVLQT1* causes the Jervell and Lange-Nielsen cardioauditory syndrome. *Nature Genet.* **15**, 186–189 (1997).
- Schulze-Bahr, E. et al. *KCNE1* mutations cause Jervell and Lange-Nielsen syndrome. *Nature Genet.* **17**, 267–268 (1997).
- Dedek, K. & Waldegger, S. Colocalization of KCNQ1/KCNE channel subunits in the mouse gastrointestinal tract. *Pflügers Arch. Eur. J. Physiol.* **442**, 896–902 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank S. Alper for the AE1 antibody, M. Knepper for the aquaporin-2 antibody, M. Knipper for advice on inner ear immunohistochemistry, and J. Enderich and M. Kolster for technical assistance. R.E. is a recipient of a Marie Curie Human Potential Fellowship of the European Union, and F.H. is a Heisenberg scholar of the Deutsche Forschungsgemeinschaft (DFG). This work was supported by grants from the DFG, the Fonds der Chemischen Industrie, and the Prix Louis Jeantet de Médecine to T.J.J., and from the Federal State of Baden-Württemberg to F.H.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.J.J. (e-mail: jentsch@zmn.uni-hamburg.de).

Isochorismate synthase is required to synthesize salicylic acid for plant defence

Mary C. Wildermuth*, Julia Dewdney*, Gang Wu†‡ & Frederick M. Ausubel*

* Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115; and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

† Michtom School of Computer Science, Brandeis University, Waltham, Massachusetts 02254, USA

Salicylic acid (SA) mediates plant defences against pathogens, accumulating in both infected and distal leaves in response to pathogen attack^{1–5}. Pathogenesis-related gene expression and the synthesis of defensive compounds associated with both local and systemic acquired resistance (LAR and SAR) in plants require SA. In *Arabidopsis*, exogenous application of SA suffices to establish SAR, resulting in enhanced resistance to a variety of pathogens. However, despite its importance in plant defence against pathogens, SA biosynthesis is not well defined. Previous work has suggested that plants synthesize SA from phenylalanine^{6–10}; however, SA could still be produced when this pathway was inhibited^{6,8}, and the specific activity of radiolabelled SA in feeding experiments was often lower than expected^{7,8}. Some bacteria such as *Pseudomonas aeruginosa* synthesize SA using isochorismate synthase (ICS) and pyruvate lyase¹¹. Here we show, by cloning and characterizing an *Arabidopsis* defence-related gene (*SID2*) defined by mutation, that SA is synthesized from chorismate by means of ICS, and that SA made by this pathway is required for LAR and SAR responses.

In *Arabidopsis*, two putative ICS genes, *ICS1* (GenBank accession numbers AC011765 and AC008263) and *ICS2* (AC011809), are located at the bottom and top of chromosome I, respectively. To test the hypothesis that plants use ICS to synthesize SA, we first examined whether pathogen treatment of *Arabidopsis* induces either of these ICS genes. Expression of *ICS1* occurred in leaves infected with the fungal biotroph *Erysiphe orontii* or the bacterial necrotroph *Pseudomonas syringae* pv. *maculicola*. In addition, when *Arabidopsis* plants were inoculated with an avirulent strain of the *Pseudomonas* pathovar *maculicola*, resulting in SA accumulation and SAR in systemic leaves, *ICS1* was systemically induced (data not shown). The timing of *ICS1* expression is similar to that of SA accumulation for these pathogen treatments and is correlated with expression of the pathogenesis-related gene *PR1*, a molecular marker of SAR^{12,13} (see Fig. 1). We detected no *ICS2* transcript in either infected or uninfected *Arabidopsis* leaves (Fig. 1).

If the product of the pathogen-inducible *ICS1* gene is involved in defence-related SA biosynthesis, an *ics1* mutant should exhibit reduced SA accumulation in response to pathogens, reduced pathogenesis-related gene expression, and enhanced susceptibility to pathogens. Two allelic *Arabidopsis* mutants, *sid2-1* (ref. 14) and *sid2-2/eds16-1* (ref. 15), exhibit these phenotypes and have been mapped to the bottom of chromosome I, near the *ICS1* locus. To ascertain whether the *sid2* mutants contain mutations in *ICS1*, we first examined the fast-neutron-generated *sid2-2* mutant. Fast neutron mutagenesis typically causes deletions resulting in loss of transcription of the affected gene. We did not detect *ICS1* transcript in *Erysiphe*-infected leaves of *sid2-2* (Fig. 2a). Moreover, amplification of *sid2-2* genomic DNA using a series of *ICS1*-specific primers

resulted in aberrant polymerase chain reaction (PCR) products. Specifically, altered products using primers in exon IX of *ICS1* suggested that a region of at least 50 nucleotides was affected (data not shown). We confirmed the presence of a significant deletion/rearrangement in exon IX by DNA blot analysis (Fig. 2b, c). To determine whether *sid2-1* (a mutant generated by treatment with ethylmethane sulphonate, EMS) contains a mutation in *ICS1*, we sequenced *ICS1* genomic DNA from *sid2-1* and wild-type plants. *sid2-1* contains a single base-pair (bp) mutation resulting in a stop codon in exon IX (Fig. 2c, d). Consequently, the mutations in both *sid2* alleles should disrupt the highly conserved chorismate-binding domain (see below).

Comparison of *Arabidopsis ICS1* with plant ICS messenger RNA sequences (AJ006065 and AF078080) revealed that *Arabidopsis ICS1* probably contains an additional amino-terminal extension not annotated for *Arabidopsis ICS1*. We therefore sequenced *Arabidopsis ICS1* complementary DNA, isolated using rapid amplification of cDNA ends (RACE)-PCR, and confirmed that an extra N-terminal extension of 66 amino acids is present in *ICS1* (Fig. 2c). This extension, absent from bacterial ICS sequences, is characteristic of chloroplast transit sequences and contains a putative cleavage site. The carboxy-terminal region of *Arabidopsis ICS1* contains the highly conserved chorismate-binding domain (Fig. 2d). Within this domain, *ICS1* contains residues essential for activity of another chorismate-binding enzyme, anthranilate synthase, except that *ICS1* has an Ala rather than a Thr residue at position 472, consistent with other ICS sequences¹⁶. Overall, the *Arabidopsis ICS1* sequence is 57% identical to ICS from *Catharanthus roseus* (Madagascar periwinkle), and greater than 20% identical to bacterial isochorismate synthases (Fig. 2e). Biochemical activity of these ICS gene products has been confirmed^{16,11}.

In *Arabidopsis*, addition of SA induces SAR and the removal of SA by the bacterial transgene *nahG* abolishes SAR^{1,2}. In *sid2* mutants, total SA accumulation is about 5–10% of wild-type levels in response to either the virulent biotroph *Erysiphe* or avirulent strains of *Pseudomonas syringae*^{14,15}, and is similar in magnitude to that of uninfected plants (Fig. 3a). We also assessed levels of pathogenesis-related protein 1 (*PR1*) because *PR1* is a SAR molecular marker downstream of SA accumulation—*PR1* is induced at very low levels in *sid2* mutants, approximately 1–10% of wild-type levels^{14,15} (Fig. 3a; see also Supplementary Information Fig. 5). Consistent with these observations, *sid2* mutants do not exhibit SAR in systemic leaves when infected with an avirulent pathogen¹⁴, or produce SAR-like responses in leaves infected with the fungal biotroph *Erysiphe*¹⁵. These results suggest that *Arabidopsis* SAR defence responses require SA synthesized using *ICS1*.

To investigate the role of *ICS1* in SAR, we examined *ICS1*

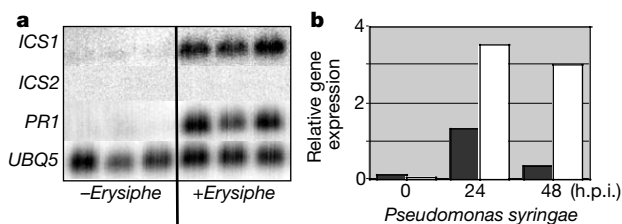


Figure 1 *ICS1* is induced in response to pathogens. RNA blot analysis of total RNA from wild-type (Columbia) *Arabidopsis* leaves probed with *ICS1*, *ICS2*, *PR1* or the loading control *UBQ5*. **a**, RNA extracted from *Erysiphe*-infected or control leaves. Triplicate samples are shown. **b**, RNA extracted from leaves inoculated with a suspension (optical density at 600 nm of 0.002) of *Pseudomonas* pathovar *maculicola* strain ES4326 at 0-, 24- and 48-h post inoculation (h.p.i.). *ICS1* (black) or *PR1* (white) normalized to *UBQ5*. *ICS1/UBQ5* values were multiplied by 100. *ICS1* probe has less than 0.1 times the activity of the *PR1* probe. No *ICS2* signal was detected. Average of duplicate samples is shown. Experiments in **a** and **b** were repeated with similar results.

‡ Present address: Strong Systems, 1077 Lexington Street, Waltham, Massachusetts 02452, USA.

expression in *Arabidopsis* mutants exhibiting altered SAR responses. Genetic analyses have identified NPR1/NIM1 (refs 17–20) as a requisite regulator of SAR that acts downstream of SA. NPR1/NIM1 interacts with transcription factors to alter the expression of genes (such as *PR1*) in response to pathogen attack²¹. Figure 3b indicates that *ICS1* acts upstream of NPR1/NIM1, as *ICS1* mRNA is still expressed in the *npr1* mutant, consistent with SA accumulation through *ICS1* acting upstream of NPR1. Our data support a role for NPR1 as a negative feedback regulator of *ICS1* expression and SA accumulation. This role for NPR1 was first proposed when elevated SA levels were observed in infected leaves of *npr1/nim1* plants¹⁸. We found that *ICS1* expression was also elevated in infected leaves of *npr1* plants (Fig. 3b and Supplementary Information Fig. 5). We wanted also to determine whether SA regulates the expression of *ICS1.nahG* transgenic plants convert SA to catechol; they do not accumulate SA, they express little *PR1* in response to pathogen, and they do not exhibit SAR^{1,2}. Expression of *ICS1* was not notably altered in *nahG* transgenic plants (Fig. 3b and Supplementary Information Fig. 5). The simplest interpretation of this result is that SA alone is not required for the induction or repression of *ICS1* expression and hence there is no direct autoregulation. Finally, to support our conclusion that SA synthesized through *ICS1* mediates SAR, we examined *ICS1* (and *PR1*) expression in the *Arabidopsis cpr* mutants (*cpr1*, *cpr5* and *cpr6*; refs 22, 23 and 24, respectively), which exhibit constitutively elevated SA levels, *PR1* expression and SAR. We found that *ICS1* was constitutively expressed in the *cpr* mutants (Fig. 3b and Supplementary Information Fig. 5). Figure 3c shows the placement of *ICS1* in the induction of SAR in *Arabidopsis* on the basis of these findings.

Our results indicate that SA synthesized through *ICS1* acts upstream of NPR1/NIM1 in the induction of *PR1* and SAR. Therefore, we analysed the *ICS1* promoter for known *cis*-acting regulatory elements associated with defence responses in plants (Fig. 3d and Supplementary Information Table 1), as this might elucidate regulatory elements and their cognate transcription factors that act upstream of SA accumulation. The *ICS1* promoter was enriched in W-box elements, as are promoters of genes, such as *PR1* (ref. 13), that are downstream of SA. W boxes are recognized by WRKY plant-specific transcription factors that regulate pathogen and stress responses²⁵. One potential explanation consistent with an enrichment of W-box elements in both the promoters of *ICS1* and downstream genes (for example, *PR1*) is that different WRKY factors regulate these responses. Reports of a variety of early- and late-acting WRKY factors activated by SA, elicitors and/or pathogens²⁵ support this view. Alternatively, shared WRKY factor(s) may suppress both *ICS1* and *PR1* expression. A W-box element in a region that negatively regulates *PR1* expression in response to a SA analogue²⁶ (linker scanning region 4, LS4) supports this explanation. Unlike the *PR1* promoter, the *ICS1* promoter does not contain the *cis*-acting regulatory elements in LS7 (bZip motif) or LS10 (NF- κ B motif) required for the induction of *PR1* by SA²⁶. This is consistent with our observation that removal of SA does not significantly affect expression of *ICS1* (Fig. 3b and Supplementary Information Fig. 5), and supports the hypothesis that a systemic signal other than SA is required for *ICS1* induction, SA accumulation and SAR¹.

The *ICS1* promoter also contains a Myb-binding site (MBSII) at a position (approximately –500 bp) similar to that of the other

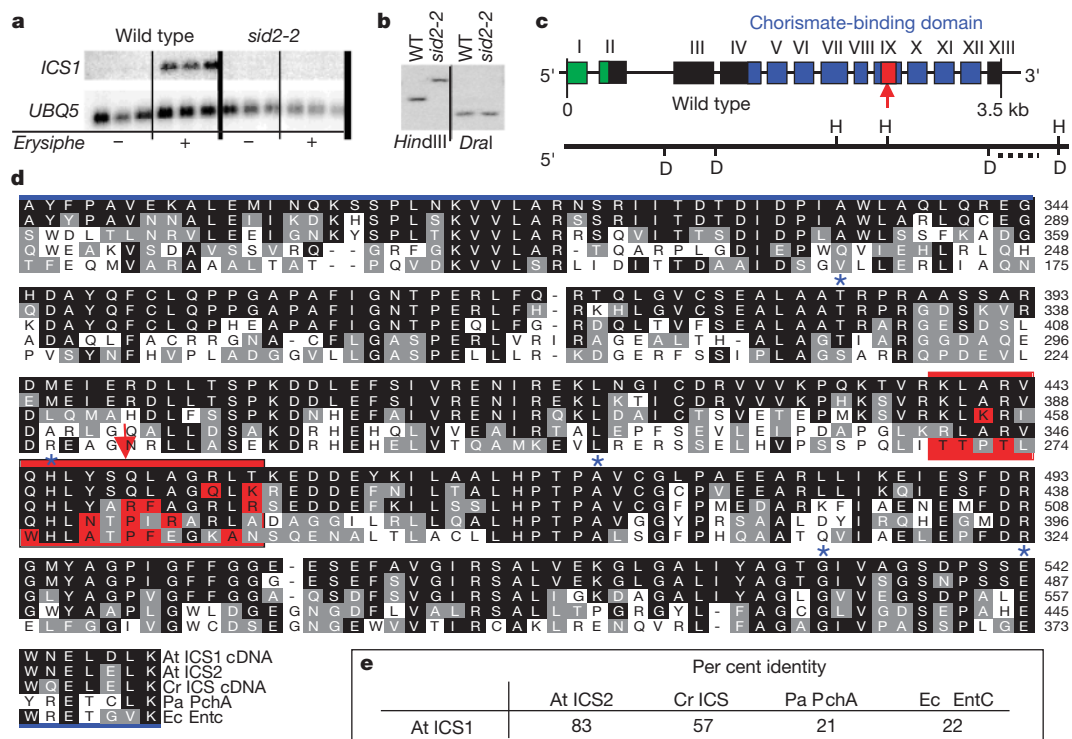


Figure 2 *SID2* encodes *ICS1*. **a**, Northern blot analysis of total RNA from *Erysiphe*-infected and uninfected leaves of *sid2-2* and wild-type *Arabidopsis* probed with *ICS1* or *UBQ5*. Triplicate samples are shown. **b**, Southern blot analysis of genomic DNA extracted from *sid2-2* or wild-type *Arabidopsis*, digested with *HindIII* or *DraI*. **c**, Schematic representation of *ICS1*. Putative plastid transit sequence (green), chorismate-binding domain (blue), and location of mutations in *sid2-2* (red box) and *sid2-1* (red arrow) are shown. Cleavage sites *HindIII* (H) and *DraI* (D) are indicated. The probe used in **b** is indicated by a dashed line. **d**, Clustal alignment of chorismate-binding domain (ProDom number P000779) of

Arabidopsis ICS1 with selected isochorismate synthases. Identical (black), conserved (shaded) and residues deemed critical for ICS activity (blue asterisk)¹⁶ are shown. T, residue 384; H, residue 445; A, residue 472; G, residue 532; E, residue 542. Location of mutations in *sid2-2* (red box) and *sid2-1* (red arrow) are shown. **e**, Overall amino-acid sequence comparison of *Arabidopsis ICS1* and selected isochorismate synthases. At *ICS1*, *Arabidopsis thaliana* cDNA (AY056055); At *ICS2*, *A. thaliana* (AAF27094); Cr *ICS*, *C. roseus* cDNA (CAA06837); Pa *PchA*, *P. aeruginosa* (CAA57969); Ec *EntC*, *Escherichia coli* (AAA16100).

inducible chorismate-using enzymes, chorismate mutase, *CM1*, (Fig. 3d and Supplementary Information Table 1), and anthranilate synthase, *ASA1* (data not shown). Myb factors have a role in SA-mediated plant defences (for example, Myb1 (ref. 27)) and in the regulation of secondary metabolism including modulation of pathogen-inducible *ASA1* expression²⁸. As the chorismate-using enzymes are critical control points of pathways resulting in many defensive compounds, it is likely that *ICS1* expression is regulated by a Myb factor(s).

We have described a SA biosynthetic pathway in *Arabidopsis* that uses isochorismate synthase to produce SA from chorismate, as do

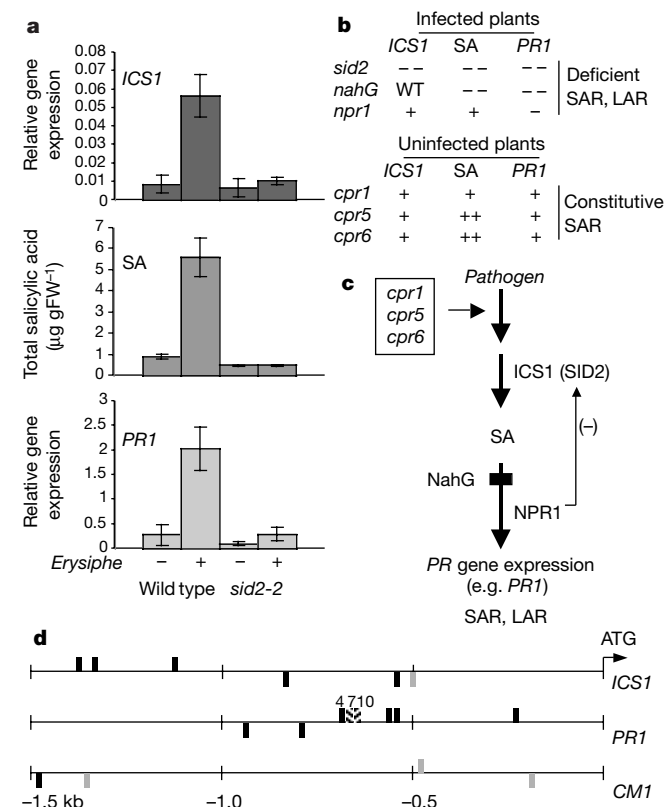


Figure 3 *ICS1* mediates SA-dependent LAR and SAR defence responses. **a**, *ICS1*, total SA levels and *PR1* gene expression in wild-type and *sid2-2* plants infected or not infected with *Erysiphe*. The average of triplicate samples is shown. Expression levels were normalized to the loading control *UBQ5*. gFW, gram fresh weight. **b**, Relative total SA levels and *ICS1* and *PR1* gene expression levels in mutant or transgenic *Arabidopsis* compared to wild-type plants. Relative values are indicated as <0.1× (–), <0.5× (–), wild type (WT), >2× (+) and >10× (++) compared with wild type. *ICS1* and *PR1* expression were assessed by northern blot analysis of total RNA infected with *Erysiphe*. Ranges represent findings from two–four separate experiments. (See Supplementary Information Fig. 5 for northern blots.) Total (free and sugar-conjugated) SA levels are compiled from Fig. 2a and published work^{14,15,18,22–24}. **c**, Proposed placement of *ICS1* in SAR induction on the basis of results tabulated in **b**. The SA-dependent NPR1-independent component of LAR/SAR is not explicitly shown. *PR1* mRNA levels are lower in leaves of *Erysiphe*-infected *sid2* than in *npr1* mutants. Therefore, *ICS1* probably participates in SA-mediated NPR1-dependent and NPR1-independent pathways. **d**, Schematic of cis-acting DNA regulatory elements 1.5-kb upstream of translation start of *ICS1*, compared with *PR1* and *CM1* (chorismate mutase). Black box, W-box core (TTGAC) or extended core (TTGAC(C/T))²⁵; grey box, Myb-binding site (MBS) II consensus (G(G/T)T(A/T)G(G/T)T)²⁷. The hatched box indicates SA analogue-inducible elements in linker scanning regions LS10 (NF-κB; GGACTTTTC) or LS7 (bZIP; ACGTCA) of the *PR1* promoter²⁶. The LS4 region (containing a W box) negatively regulates SA analogue-inducible expression of *PR1*. Boxes above the line represent elements found on the forward (becomes the coding) strand. Further details are provided in Supplementary Information Table 1.

some bacteria (Fig. 4). The *ICS1* gene contains a putative plastid transit sequence and cleavage site consistent with its use of plastid-synthesized chorismate as a substrate. Plastid-localized synthesis of SA mediated by ICS is consistent with observations that many nuclear-encoded, plastid-localized metabolic pathways derive evolutionarily from prokaryotic endosymbionts. A prokaryotic origin of *ICS1* is supported by the presence of an *ICS* gene in the chloroplast genome of the red algae *Cyanidium caldarium* (GenBank accession number NC_001840). It seems probable that other plant species also use the ICS pathway to synthesize SA, as plastid-localized periwinkle *ICS* has been identified (AJ006065)¹⁶. Moreover, expressed sequence tags for isochorismate synthase have been annotated for soybean (AW596452 and AW759689) and identified by homology and conservation of critical residues (see Fig. 2d) in other plant species, including wild tomato (AW398687).

An unresolved issue is whether SA is synthesized through the phenylalanine ammonium lyase (PAL) pathway, shown in Fig. 4, as well as through ICS. Our evidence suggests that SA synthesized through *ICS1* has an important role in plant defence against pathogens and specifically, that it is required for *PR1* gene expression and LAR and SAR defence responses. However, SA also potentiates plant cell death in response to particular pathogens or fungal elicitors¹. The importance of PAL in this type of cell death has been reported previously^{5,6}. Furthermore, *sid2* mutants infected with necrotizing pathogens still exhibit cell death^{14,15}. Therefore, the SA that potentiates plant cell death is probably synthesized through PAL. Notably, previous work on SA biosynthesis in plants focused on infected leaves undergoing cell death^{6–10}, and thus may have been biased towards elucidation of the PAL pathway. A related issue is the source of the low levels of SA present in uninfected leaves. Because these low levels are still present in *sid2* mutants, it seems that *ICS1* is not responsible for this SA production. Perhaps *ICS2* is involved in synthesis of this low level, constitutive SA. This would be analogous to the roles of the two anthranilate synthase genes in *Arabidopsis*—*ASA1* is highly induced in response to

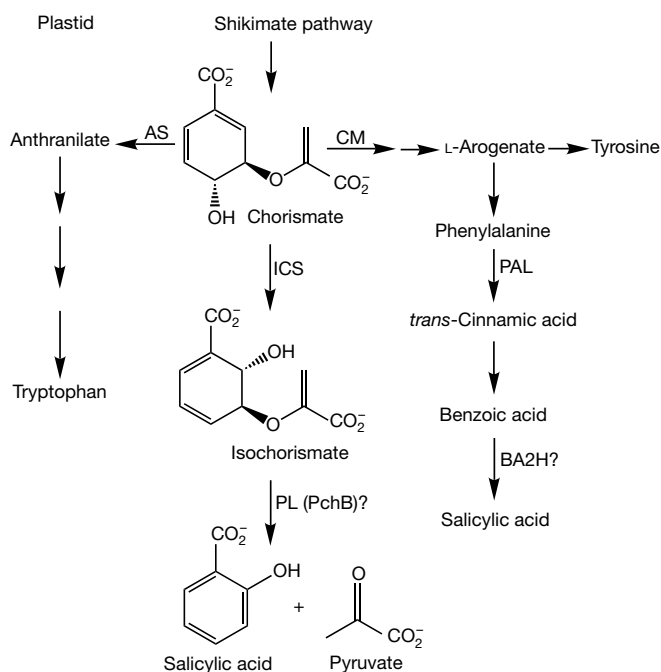


Figure 4 Proposed salicylic acid biosynthetic pathways in plants. Proposed pathway for SA biosynthesis through isochorismate synthase (ICS). In bacteria, this pathway also requires pyruvate lyase (PL; PchB in *P. aeruginosa*)¹¹. The previously described pathway of SA biosynthesis from PAL and benzoic acid-2 hydroxylase (BA2H) is also shown. AS, anthranilate synthase; CM, chorismate mutase.

pathogens and wounding, whereas ASA2 is constitutively expressed at low levels²⁹. Alternatively, there is evidence that the PAL pathway may be operational under these conditions^{7,8}. There are at least three PAL genes in *Arabidopsis*, one of which could function to produce these low levels of SA.

A mechanistic understanding of the functions of SA in plant defence requires knowledge not only of where, when, and how much SA is made, but how it is synthesized. Our work now allows for constructs and genetic mosaics to be created which will allow for the careful examination of the complex role of SA in plant defence, and for the identification of signals and regulators of SAR that act upstream of SA. □

Methods

RNA analysis

Arabidopsis mutants and transgenics are of the Columbia ecotype. The *nahG* transgenic *Arabidopsis* line and the *npri-1*, *cpr1-1*, *cpr5-1* and *cpr6-1* mutant alleles have been described^{17,22–24,30}. At 4–4.5 weeks of age, plants were infected with *P. syringae* pv. *maculicola* strain ES4326 or *E. orontii* isolate MGH. *Erysiphe*-infected leaves were collected at 7 days post inoculation, and RNA extractions and blot analysis (including synthesis of *PR1* and *UBQ5* probes) were performed as described¹⁵. We used a Molecular Dynamics PhosphorImager to quantify the levels of hybridization. *ICS1* and *PR1* hybridization levels were normalized to those obtained for the loading control *UBQ5*. *ICS1*- and *ICS2*-specific primers corresponded to the 3'-untranslated region. Forward and reverse primers used to make the *ICS* probe templates and probes were: *ICS1*, forward 5'-GGGGATAAGGGGTTTCACAAATA-3' and reverse 5'-CTGCCCTAGTTACAAACCCGAAAAG-3'; *ICS2*, forward 5'-TGTGTTTGGTCATTGGTGT-3' and reverse 5'-TCATAGAGTCATAGTCGCTTCA-3'. One, five and ten picograms of *ICS1* and *ICS2* probe templates were spotted onto nylon membranes before crosslinking as controls for successful hybridization and to assess cross-hybridization (negligible).

Salicylic acid

We extracted and quantified total salicylic acid (free and sugar-conjugated) by HPLC as described¹⁵. Recovery was determined for each sample by spiking with *o*-anisic acid before leaf extraction.

sid2-2 and *sid2-1* mutations

Genomic DNA was extracted from the fast-neutron-generated mutant *sid2-2* and from wild-type *Arabidopsis*, and digested with *HindIII* or *DraI*. DNA blot analysis was performed with the probe shown in Fig. 2c using standard procedures (CPMB). The unaltered *DraI* cleavage pattern indicates that this deletion/rearrangement does not extend into the next gene. Additional restriction enzymes and probes gave results consistent with the alteration in exon IX (see Fig. 2c, d).

The *ICS1* sequence from the *sid2-1* mutant generated with EMS, and wild-type (Columbia) genomic DNA was amplified by PCR (using Pfu Taq), subcloned and sequenced using standard procedures (CPMB). Multiple overlapping sequences of *sid2-1* contained a mutation (C→T in the coding strand) in exon IX. The position of the EMS-generated mutation in *sid2-1* results in a stop codon (TAA) at residue 449 instead of glutamine. The wild-type (Columbia) sequence matched the GenBank genomic sequence. This experiment was repeated with the same result.

Isolation of cDNA, annotation and sequence analysis

ICS1 cDNA sequence was isolated using RACE-PCR (Clontech SMART RACE cDNA Amplification Kit). Predicted chloroplast localization and transit sequence cleavage site were determined by analysis with ChloroP (<http://www.cbs.dtu.dk/services/ChloroP>). We assessed comparisons of overall sequence identity using DNASTar MegAlign Clustal analysis program. Conserved residues are within three distance units (PAM250 matrix). The N-terminal extensions (putative plastid transit sequences) of plant isochorismate synthases were not included for this analysis.

Received 14 May; accepted 27 September 2001.

- Dempsey, D. M. A., Shah, J. & Klessig, D. F. Salicylic acid and disease resistance in plants. *Crit. Rev. Plant Sci.* **18**, 547–575 (1999).
- Ryals, J. A. *et al.* Systemic acquired resistance. *Plant Cell* **8**, 1809–1819 (1996).
- Malamy, J., Carr, J. P., Klessig, D. F. & Raskin, I. Salicylic acid: a likely endogenous signal in the resistance response of tobacco to viral infection. *Science* **250**, 1002–1004 (1990).
- Metraux, J.-P. *et al.* Increase in salicylic acid at the onset of systemic acquired resistance in cucumber. *Science* **250**, 1004–1006 (1990).

- Dorey, S. *et al.* Spatial and temporal induction of cell death, defense genes, and accumulation of salicylic acid in tobacco leaves reacting hypersensitively to a fungal glycoprotein elicitor. *Mol. Plant Microbe Interact.* **10**, 646–655 (1997).
- Mauch-Mani, B. & Slusarenko, A. J. Production of salicylic acid precursors is a major function of phenylalanine ammonia-lyase in the resistance of *Arabidopsis* to *Peronospora parasitica*. *Plant Cell* **8**, 203–212 (1996).
- Yalpani, N., Leon, J., Lawton, M. A. & Raskin, I. Pathway of salicylic acid biosynthesis in healthy and virus-inoculated tobacco. *Plant Physiol.* **103**, 315–321 (1993).
- Coquoz, J.-L., Buchala, A. & Metraux, J.-P. The biosynthesis of salicylic acid in potato plants. *Plant Physiol.* **117**, 1095–1101 (1998).
- Ribnicky, D. M., Shulaev, V. & Raskin, I. Intermediates of salicylic acid biosynthesis in tobacco. *Plant Physiol.* **118**, 565–572 (1998).
- Leon, J., Shulaev, V., Yalpani, N., Lawton, M. A. & Raskin, I. Benzoic acid 2-hydroxylase, a soluble oxygenase from tobacco, catalyzes salicylic acid biosynthesis. *Proc. Natl Acad. Sci. USA* **92**, 10413–10417 (1995).
- Serino, L. *et al.* Structural genes for salicylate biosynthesis from chorismate in *Pseudomonas aeruginosa*. *Mol. Gen. Genet.* **249**, 217–228 (1995).
- Uknes, S. *et al.* Biological induction of systemic acquired resistance in *Arabidopsis*. *Mol. Plant Microbe Interact.* **6**, 692–698 (1993).
- Maleck, K. *et al.* The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nature Genet.* **26**, 403–410 (2000).
- Nawrath, C. & Metraux, J.-P. Salicylic acid induction-deficient mutants of *Arabidopsis* express *PR-2* and *PR-5* and accumulate high levels of camalexin after pathogen inoculation. *Plant Cell* **11**, 1393–1404 (1999).
- Dewdney, J. *et al.* Three unique mutants of *Arabidopsis* identify *eds* loci required for limiting growth of a biotrophic fungal pathogen. *Plant J.* **24**, 205–218 (2000).
- van Tegelen, L. J. P., Moreno, P. R. H., Croes, A. E., Verpoorte, R. & Wullems, G. J. Purification and cDNA cloning of isochorismate synthase from elicited cell cultures of *Catharanthus roseus*. *Plant Physiol.* **119**, 705–712 (1999).
- Cao, H., Bowling, S. A., Gordon, A. S. & Dong, X. Characterization of an *Arabidopsis* mutant that is nonresponsive to inducers of systemic acquired resistance. *Plant Cell* **6**, 1583–1592 (1994).
- Delaney, T., Friedrich, L. & Ryals, J. *Arabidopsis* signal transduction mutant defective in chemically and biologically induced disease resistance. *Proc. Natl Acad. Sci. USA* **92**, 6602–6606 (1995).
- Cao, H., Glazebrook, J., Clarke, J. D., Volko, S. & Dong, X. The *Arabidopsis* NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. *Cell* **88**, 57–63 (1997).
- Ryals, J. *et al.* The *Arabidopsis* NIM1 protein shows homology to the mammalian transcription factor inhibitor 1 kappa B. *Plant Cell* **9**, 425–439 (1997).
- Despres, J., DeLong, C., Glaze, S., Liu, E. & Fobert, P. R. The *Arabidopsis* NPR1/NIM1 protein enhances the DNA binding activity of a subgroup of the TGA family of bZIP transcription factors. *Plant Cell* **12**, 279–290 (2000).
- Bowling, S. A. *et al.* A mutation in *Arabidopsis* that leads to constitutive expression of systemic acquired resistance. *Plant Cell* **6**, 1845–1857 (1994).
- Bowling, S. A., Clarke, J. D., Liu, Y., Klessig, D. F. & Dong, X. The *cpr5* mutant of *Arabidopsis* expresses both NPR1-dependent and NPR1-independent resistance. *Plant Cell* **9**, 1573–1584 (1997).
- Clarke, J. D., Liu, Y., Klessig, D. F. & Dong, X. Uncoupling PR gene expression from NPR1 and bacterial resistance: characterization of the dominant *Arabidopsis* *cpr6-1* mutant. *Plant Cell* **10**, 557–569 (1998).
- Eulgem, T., Rushton, P. J., Robatzek, S. & Somssich, I. E. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* **5**, 199–206 (2000).
- Lebel, E. *et al.* Functional analysis of regulatory sequences controlling PR-1 gene expression in *Arabidopsis*. *Plant J.* **16**, 223–233 (1998).
- Yang, Y. & Klessig, D. F. Isolation and characterization of a tobacco mosaic virus-inducible myb oncogene homolog from tobacco. *Proc. Natl Acad. Sci. USA* **93**, 14972–14977 (1996).
- Bender, J. & Fink, G. R. A Myb homologue, ATR1, activates tryptophan gene expression in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **95**, 5655–5660 (1998).
- Niyogi, K. K. & Fink, G. R. Two anthranilate synthase genes in *Arabidopsis*: defense-related regulation of tryptophan pathway. *Plant Cell* **4**, 721–733 (1992).
- Reuber, T. L. *et al.* Correlation of defense gene induction defects with powdery mildew susceptibility in *Arabidopsis* enhanced disease susceptibility mutants. *Plant J.* **16**, 473–485 (1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank S. Volk for RNA from leaves infected with *Pseudomonas* pathovar *maculicola* strain ES4326, C. Nawrath for the *sid2-1* seeds, and D. Lee, J. Stone, J. Dangel and E. Meyerowitz for comments on drafts of the manuscript. This work was supported by a National Institutes of Health grant to F.M.A. and a National Research Initiative Competitive Grants Program/USDA postdoctoral fellowship to M.C.W.

Correspondence and requests for materials should be addressed to F.M.A. (e-mail: ausubel@molbio.mgh.harvard.edu). The sequence for *ICS1* cDNA is available at GenBank under accession number AY056500.

Four assays and a fusion

Microarrays for the novice, expressive messengers and a breaking story.

ResGen GeneFilters

Invitrogen www.resgen.com
All-in-one microarray kit

This starter pack for microarray screening analysis contains two microarrays for experimentation, a test filter to optimize hybridization, a probe labelling kit to radiolabel and purify 10 high-quality cDNA probes, hybridization buffer and Pathways software for comprehensive differential gene expression analysis. Invitrogen says that buying the kit in place of individual components provides significant cost savings.

Microgenomics system

Arcturus www.arctur.com
mRNA for gene expression studies

An optimized, experimentally validated instrument and reagent system for studying gene expression in small, pure cell populations is now available. The system consists of a HistoGene LCM frozen section staining kit, a PixCell II laser capture microdissection system, a PicoPure RNA isolation kit and a RiboAmp RNA amplification kit to procure and amplify high-quality mRNA for cell-specific gene expression studies.

Human Ku 70/80

Trevigen www.trevigen.com
Complex repair for DNA breaks

Human Ku 70/80 complex binds at double-stranded DNA breaks and recruits DNA-PK, RAD50-Mre11-Xrs2, and XRCC4/DNA ligase IV to the damaged site to facilitate non-homologous end joining (NHEJ). It also participates in NHEJ during the V(D)J recombination of immunoglobulin genes needed to generate the large antibody repertoire. In addition, the complex restores activity to an immunodepleted cell free DNA repair assay, binds to the natural ends of chromosomes within telomeres and may play a role in the regulation of telomere length. Auto-antibodies to Ku have been found in the sera of patients with systemic lupus mixed connective tissue disease, scleroderma, polymyositis and Grave's disease. Applications include studies of double-stranded break repair and non-homologous end joining of DNA.



Screen test: Pierce's new kit for assay and purification of *E. coli* 6 × His fusion proteins.

Lyse-N-Assay

Pierce www.piercenet.com
Versatile system for 6 × His fusion proteins

Many or just a few 6 × His fusion proteins from *E. coli* can be screened at the same time with this kit. It includes either clear or white 96-well HisGrab high binding capacity plates for automation and B-PER II bacterial protein extraction reagent for cell lysis. Screening, purification and further downstream applications can be performed in the same plate.

Mouse IFN-α ELISA

PBL Biomedical Laboratories www.interferonsource.com
A must for detection of mouse interferon

Created specifically for measuring mouse interferon alpha (Mu-IFN-α), this kit detects Mu-IFN-αA, α1, α4, α5, α6 and α9 in serum or media using a sandwich immunoassay. It shows no cross-sensitivity with mouse IFN-β, mouse IFN-γ or human IFN-α. The kit includes all required reagents plus a pre-coated, 96-well microtitre breakaway strip plate.

NEFA C

Wako Diagnostics www.wakousa.com
In vitro colorimetric assay

Extraction procedures are unnecessary with this kit for quantitative determination of non-esterified (or free) fatty acids in serum or plasma (non-heparinized). The kit uses *in vitro* enzymatic colorimetric methodology that can be automated with existing clinical chemistry instrumentation. The method relies on acylation of co-enzyme A (CoA) by fatty acids in the presence of added acyl-CoA synthetase (ACS). Linear range of the assay is 0–2.0 mmol per litre

Inoculating turntable

Bel-Art Products www.bel-art.com
Even distribution of bacterial colonies

This instrument produces nearly concentric circles of evenly distributed bacterial colonies

on 60, 95 and 100 mm diameter Petri dishes. The turntable is built for simplicity and low weight, and is chemical-resistant and portable. It is hand-operated with ball-bearings ensuring a smooth, even motion for consistent distribution of cultures. Extended spin time speeds completion of the inoculation.

Versalinx

Prolinx www.prolinxinc.com
High-capacity beads and protein release

Versalinx chemistry is based on two small synthetic molecules, phenylboronic acid (PBA) and salicylhydroxamic acid (SHA). SHA-modified SPM-HC Beads bind PBA-modified proteins with low nonspecific binding, while the Versalinx Protein Releasing Reagent allows recovery of PBA-protein under mild conditions, if desired. Attributes allow for the incorporation into traditional or high-throughput proteomics methods, including affinity purification and analytical assays.

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENT

BACs • YACs • PACs cDNAs

We provide clones and colony membranes for a number of human, mouse and rat libraries as well as a custom screening service for the libraries.

ResGen™
Invitrogen Corporation

U.S. or Canada 800-533-4363
FAX 256-536-9016

Check out our web site at
www.resgen.com

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:

Leonie Welss (4954)

European Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy:

Nevin Bayoumi (4978)

Scandinavia:

Sille Opstrup (4994)

Spain/ Portugal:

Leonie Welss (4954)

Production Manager:

Billie Franklin

To send films and materials use

London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

Tel + 49 89 54 90 57 11/2

Fax + 49 89 54 90 57 20

e-mails: p.phelan@nature.com

k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Corissa Salzman

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakakatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.cowan@naturejpn.com

Japan Manager:

Kate Cowan

naturejobs

A shift to the centre

Years ago, *The New Yorker* magazine published an illustrated map of the United States that soon became a classic. It showed the East and West coasts dominated by tall buildings, with the centre of the country represented as a disproportionately tiny wasteland. This misconception persists. Recently, New York mayor-elect Michael Bloomberg quipped that any business that moved from New York City "might as well be in Iowa in the cornfields".

In fact, the former fields of Iowa and other midwestern states are providing fertile ground for science and technology. Michigan is pouring millions of dollars into a 'life sciences corridor' (see Spotlight, inside). Washington University in St Louis, Missouri, continues to be a major contributor to the human genome project. And the Stowers Institute in Kansas City, Missouri, is recruiting internationally to fill its new campus (see Movers, back page).

Midwestern research institutions, like their coastal cousins, also create synergistic relationships with industry. For example, in 1925 a group of alumni from the University of Wisconsin-Madison established a licensing body to commercialize vitamin D, discoveries made by Professor Harry Steenbock. It has since obtained over 1,000 patents, including one for immortalized human embryonic stem-cell lines. As well as aiding the creation of spin-off companies, such arrangements mean money gets ploughed back into the university. Both outcomes result in an abundance of jobs. Madison, like other biotech hotspots around the world, tends to have a much lower unemployment rate and a higher per capita income level.

And scientists who relocate from Cambridge or Palo Alto to Madison, Missouri or Michigan will find housing prices and traffic congestion are both far more favourable — proving that the open spaces mocked by coastal defenders can actually be an advantage.

Paul Smaglik

Naturejobs editor



Contents

MOVERS

Biotech veteran is tempted out of retirement; immunologists head for the private sector; Beckman Institute gets new head; and more [Back page](#)

WWW.NATUREJOBS.COM

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

BIOTECH

After less than a year in retirement, **David Konys** has been tempted back into employment. He joined Ingenium Pharmaceuticals, a company based in Martinsried, Germany, in May and last month was appointed vice-president of corporate development. Konys, who spent 15 years with Biogen in Cambridge, Massachusetts, before retiring last year, will operate Ingenium's new subsidiary in Boston. He left Biogen after turning 50 when he realized that, with his children through college and his house paid off, he didn't really need to work any more. Konys had envisioned an idyllic existence travelling, writing and upgrading his pilot skills. But a chance encounter with a venture capitalist, who understood that someone with 15 years of biotech experience is rare, led him back. Konys anticipates retiring again in five years — maybe.

The recent move from the public sector to a private company by **Jaap Goudsmit** has encouraged a fellow immunologist to take the plunge. Goudsmit is now senior vice-president, vaccine research, for Crucell, a biotechnology company based in Leiden, the Netherlands. In January, he will be joined by **Ada Kruisbeek**, who last month accepted an offer to become the company's senior vice-president, head of oncology and inflammatory diseases.

Goudsmit is a co-founder of the International AIDS Vaccine Initiative (IAVI) and EuroVac, the European Union's AIDS vaccine effort. Before joining Crucell he was chairman of the University of Amsterdam's Research Institute for Infectious Diseases and the Institute for Science Education. He will remain affiliated with the university, and will continue his involvement in the IAVI and EuroVac. For the past 10 years, Kruisbeek has worked at the Netherlands Cancer Institute in Amsterdam, first as chair of the department of immunology, and more recently as laboratory research coordinator for all research divisions.

BIOLOGY

The Beckman Institute at the California Institute of Technology (Caltech) has a new director. Biology professor **Barbara Wold**, who has been at Caltech since 1981, succeeds founding director **Harry Gray**, who will return to full-time professorial duties after running the institute for 15 years. Wold specializes in embryonic development and regeneration in vertebrates. She will lead the institute in its development of pioneering interdisciplinary work in biology and chemistry.

The Stowers Institute for Medical Research has attracted two more biologists to its campus in Kansas City, Missouri, including a French developmental biologist who has never previously worked outside France. **Olivier Pourquié** is currently an independent research group leader at the Developmental Biology Institute of Marseille. He will join the Stowers Institute as an associate scientist next June, along with **Jennifer Gerton**, who will be an assistant scientist with the institute after she completes a postdoc at the University of California, San Francisco.

GENOMICS

The Cornell Genomics Initiative Task Force, launched in 1997, continues to expand, with 18 members appointed by August. The most recent faculty to join are **Rebecca Nelson**, associate professor of plant pathology, who was a molecular pathologist at the International Potato Center in Lima, Peru; **Teresa Gunn**, assistant professor of genetics, who comes from a Stanford postdoc; **Helene Marquis**, assistant professor of bacteriology, who was previously an assistant professor of microbiology at the University of Colorado Health Sciences Center in Denver; and **Marci Scidmore**, assistant professor of microbiology and immunology, who was a senior staff fellow at the National Institute of Allergy and Infectious Diseases Rocky Mountain Laboratories in Hamilton, Montana. By 2005, the initiative is expected to have as many as 50 members.

PHYSICS

Jeffrey Jaffe last month replaced **William Brinkman** as research vice-president at Bell Labs. Jaffe has a background in software development. Before joining Lucent in 2000, Jaffe worked at IBM for 16 years. He has also advised the US government on Internet issues, including serving in 1997 on an advisory committee to President Bill Clinton's Commission for Critical Infrastructure Protection. Brinkman, who had been at Bell Labs since 1966, except for three years at Sandia National Laboratories in Albuquerque, New Mexico, will continue to serve as an adviser.

ERRATUM

In Movers in the 15 November issue of *Naturejobs*, the photograph captioned Andrew Webb is, in fact, a picture of Danny Powell. In addition, the second paragraph in the story about Webb's move starts with the word 'Lenox', this should read 'Webb'. *Naturejobs* apologizes for any confusion.



David Konys



Ada Kruisbeek



Olivier Pourquié

CONTACTING US AT MOVERS

Please send any information on new appointments or job moves to: naturejobseditor@naturedc.com.